

Student thesis series INES nr 394

# Exploring the spatial relationship between NDVI and rainfall in the semi-arid Sahel using geographically weighted regression

**Stefanos Georganos**

---

2016  
Department of  
Physical Geography and Ecosystem Science  
Lund University  
Sölvegatan 12  
S-223 62 Lund  
Sweden



Stefanos Georganos (2016).

***Exploring the spatial relationship between NDVI and rainfall in the semi-arid Sahel using geographically weighted regression***

Master degree thesis, 30 credits in Master of *Geomatics*

Department of Physical Geography and Ecosystem Science, Lund University

Level: Master of Science (MSc)

Course duration: *January* 2016 until *June* 2016

#### Disclaimer

This document describes work undertaken as part of a program of study at the University of Lund. All views and opinions expressed herein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

# Exploring the spatial relationship between NDVI and rainfall in the semi-arid Sahel using geographically weighted regression

---

Stefanos Georganos

Master thesis, 30 credits, in *Geomatics*

**Supervisor:** Abdulhakim Abdi

Department of Physical Geography and Ecosystem Science  
Lund University, Sweden

**Supervisor:** David Tenenbaum

Department of Physical Geography and Ecosystem Science  
Lund University, Sweden

**Exam committee:**

Jonas Ardö, Department of Physical Geography and Ecosystem Science  
Lund University, Sweden

Finn Hedefalk, Department of Physical Geography and Ecosystem Science  
Lund University, Sweden



## **Abstract**

The Normalized Vegetation Difference Index (NDVI) is frequently used as a surrogate for vegetation properties and is often correlated with climatic variables such as rainfall. However, studies have shown that conventional regression models used to study the spatial relationship between NDVI and rainfall are often plagued by non-stationarity and are scale dependent. This thesis employed a spatial disaggregation modelling technique to tackle this issue – Geographically Weighted Regression (GWR) allows measured relationships to vary in space. GWR was applied in the Sahel of Africa for the growing seasons of 2002 and 2012 (June-September). The results showed that with the selection of an appropriate scale, the spatial patterns of NDVI were significantly better explained by applying the GWR (0.85 and 0.80  $R^2$ ) than Ordinary Least Squares (OLS) models (0.68 and 0.62  $R^2$ ) both in terms of predictive power, accuracy and reduced residual autocorrelation. Moreover, clear spatial clusters were formed with coefficients significantly higher or lower than those a global model would suggest. Areas near wetlands and irrigated lands displayed weak correlations while humid areas such as the Sudanian region of the Sahel produced very high correlations. Finally, the spatial relationship of rainfall and NDVI displayed temporal variations as there were differences in the local coefficients between a relatively dry and wet year. Therefore, GWR is suggested as an accurate, informative technique both for exploratory and explanatory purposes to address non-stationarity in spatially heterogeneous areas in an ecological context.

**Keywords:** geographically weighted regression, non-stationarity, scale dependency, Sahel, vegetation



# Table of Contents

1.	<b>Introduction</b> .....	1
1.1.	Background.....	1
1.2.	Objectives and Research Questions.....	3
1.3.	Study Area.....	5
1.4.	Thesis Outline.....	7
2.	<b>Background and Literature Review</b> .....	9
2.1.	Introduction.....	9
2.2.	Vegetation-rainfall relationship in the Sahel.....	9
2.3.	Geographically Weighted Regression in ecological modelling.....	10
2.4.	Criticisms of Geographically Weighted Regression.....	11
3.	<b>Data and Methodology</b> .....	13
3.1.	Data.....	13
3.1.1.	Normalized Difference Vegetation Index – NDVI.....	13
3.1.2.	AVHRR GIMMS NDVI.....	13
3.1.3.	Climate Hazard Infrared Precipitation with Stations (CHIRPS).....	15
3.2.	Methodology.....	17
3.2.1.	Geographically weighted regression.....	16
3.2.2.	Scale dependency of non-stationarity.....	19
3.2.3.	Model comparison.....	20
3.2.4.	Tests for the significance of spatial non-stationarity.....	21
3.2.5.	Testing for residual autocorrelation.....	22
3.2.6.	Data processing and software.....	22
3.2.7.	Land cover analysis.....	23
4.	<b>Results</b> .....	25
4.1.	Scale dependency.....	25
4.2.	Spatial patterns of the NDVI – rainfall relationship.....	26
4.3.	Land cover analysis.....	30
4.4.	Model Diagnostics and tests for non-stationarity.....	32
4.5.	Predicted Patterns.....	34

4.6.	Spatial autocorrelation of the residuals.....	38
5.	<b>Discussion</b> .....	43
6.	<b>Conclusions</b> .....	49
	<b>Aknowledgments</b> .....	51
	<b>References</b> .....	53
	<b>List of Publications</b> .....	60



## List of Figures

Figure 1: Map of the study area .....	5
Figure 1: Land cover map of the Sahel based on the GLC classification.....	6
Figure 1: Integrated NDVI and cumulative rainfall in the Sahel in 2002.....	14
Figure 1: Integrated NDVI and cumulative rainfall in the Sahel in 2012.....	14
Figure 1: Gaussian Spatial Weighting function for GWR .....	19
Figure 1: Selection of bandwidth size.....	20
Figure 1: Stationarity index for 2002 and 2012 at multiple scales .....	26
Figure 1: Local R2 patterns in 2002 and 2012.....	28
Figure 1: Rainfall (Beta) coefficient in the Sahel for 2002 and 2012.....	29
Figure 1: Spatial Clustering Maps of the slope coefficient in 2002 and 2012.....	30
Figure 1: Scatter plots of observed and simulated NDVI in both growing seasons a .....	36
Figure 1: Observed and Predicted spatial patterns of iNDVI in 2002 .....	37
Figure 1: Observed and Predicted spatial patterns of iNDVI in 2012 .....	38
Figure 1: Residual distribution in 2002 for OLS and GWR models.....	39
Figure 1: Residual distribution in 2012 for OLS and GWR models.....	40
Figure 1: Spatial correlograms in 2002 and 2012 .....	41

## **List of Tables**

Table 1 : Regression Correlation Coefficients and AICc land cover categories .....	31
Table 2: Model comparison between GWR and OLS models in 2002 and 2012.....	32
Table 3: Descriptive statistics for regression parameters in 2002 .....	33
Table 4: Descriptive statistics for regression parameters in 2012. ....	34

# **1. Introduction**

## **1.1 Background**

Satellite remote sensing is often used to estimate vegetation distribution and productivity at large spatial scales. The normalized difference vegetation index (NDVI) is the most widely used surrogate for large-scale assessments of vegetation greenness and has been applied in a wide range of studies (Brandt et al. 2015, Chen et al. 1998; Santos and Negri, 1997; Zhang et al 2009). The spatial distribution of remotely sensed NDVI and consequently of terrestrial vegetation, is a function of prevalent climatic conditions such as rainfall and temperature. The relationship between NDVI and rainfall is well established at various spatial and temporal scales (Davenport et al. 1993; Grist et al. 1997; Nicholson et al. 1990; Potter and Brooks 1999; Wang et al. 2001). The results of these studies, although varying, indicate that rainfall is an important predictor of the geographical distribution of vegetation in many environments, particularly in transitional zones, such as from humid to arid and semi-arid environments (Zhao et al. 2015) as found in the Sahel of Africa.

In most studies that characterize the relationship between vegetation and rainfall, NDVI is often modelled as function of rainfall using linear global models such as Ordinary Least Squares (OLS) regressions. However, the NDVI-rainfall relationship varies spatially and temporally depending on land cover, soil type, vegetation composition and structure, microclimatic conditions and human impact (Foody 2003; Li et al. 2002; Richard and Pocard 1998; Propastin et al. 2007). This phenomenon is known in regression modelling as non-stationarity. Non-stationarity can be a problem when modelling the NDVI-rainfall relationship (Eklundh 1998) especially when traditional, linear-based models are used. In a large-scale spatial analysis, it is reasonable to assume that both NDVI and rainfall will vary to some extent and therefore, making the use of global correlation models such as linear regressions not successfully capture the true relationship between the two variables. Foody (2003) discussed in detail the potential dangers and flaws of global regression modelling with remotely sensed data. The

assumption that a single set of regression parameters can accurately describe a spatially varying relationship has been highly questioned in the past two decades in landscape ecology and geography. This can potentially lead to false assumptions regarding the underlying processes affecting vegetation distribution (Gao and Li 2011). Indeed, studies have shown that NDVI has significant spatial variation when modelled with most environmental variables (Propastin 2009; Zhao et al. 2015). The term “scale effect” in ecological terms was highlighted by Foody (2003), to show that environmental relations are a function of the spatial scale they are being examined.

Global models assume that a relationship between two or more variables is stationary and that the resulting regression outputs can be well applied throughout the study area. However, due to spatial heterogeneity these parameters often fail to establish the proper nature and strength of the associations and may not apply to any location at all even if the relationship between variables is found to be strong (Osborne and Suarez 2002). Since spatial non-stationarity violates one of the basic regression assumptions – spatial independence between observations (Anselin 1988), the need for statistical models that can account for this phenomenon may be profound. Some regression models have limited inclusion of the spatial structure of the data. In general, and depending on the application, they aim to increase the prediction power and accuracy and deal with residual autocorrelation. They mostly utilize spatial lags such as simultaneous autoregressive models (SAR) (Kissling and Carl 2008), Poisson regressions (Kalogirou 2015), or error decomposition models (Bolduc et al. 1995). However, they do not deal with non-stationary data directly but rather try to “model” the error dependency (Maantay and McLafferty 2011). Consequently, the issue of facing non-stationary data remains and global models are frequently deemed incapable of dealing with it effectively. This can be problematic in ecological studies since remotely sensed data have an explicit spatial nature, often covering large surfaces with varying characteristics (Foody 2002).

This thesis attempts to tackle this commonly encountered issue by employing a local regression method to deal with non-stationarity, known as Geographically Weighted Regression (GWR). GWR is a nonparametric technique (Fotheringham et al. 2002) and although a dominant application in the fields of human geography (Cahill and Mulligan

2007; Fotheringham et al. 2001; Hu et al. 2012; Kalogirou 2003) it has only recently started to gain ground in the field of ecology (Gaughan et al. 2012; Propastin et al. 2008; Wang et al. 2005; Zhao et al. 2015). GWR allows the relationships between dependant and explanatory variables to vary over space, dealing with both non-stationarity and residual correlation in the most direct way possible. The outputs of this local method can be illuminating for descriptive purposes and can be used to detect areas of model misspecification or areas that show interesting variations that otherwise would be lost in the form of a global model (Brunsdon et al. 1998). In large-scale ecological studies, more often than not, the GWR model is shown to be better applied by standard statistical measures such as the Akaike Information Criterion (AIC) or cross validation scores (CV) in comparison to global regression models. However, that markedly depends on the nature of the dataset and the degree of non-stationarity.

The Sahel of Africa, as an ecoclimatic transition zone between arid to humid climates is known to be sensitive to environmental changes (Nicholson et al. 1990). Some of these climatic transition zones are often known to emerge with complex characteristics in regards to the relation of NDVI and climatic variables such as non-stationarity, scale dependency and spatial heterogeneity (Zhao et al. 2015). Therefore, it would be of significant interest to tackle these topics by employing GWR as a method to explore these issues and potentially improve modelling of the relationship between NDVI and rainfall in the region.

## **1.2 Objectives and Research Questions**

The aim of this thesis is to explore the NDVI – rainfall spatial relationship in the Sahel zone of Africa during the growing seasons of 2002 and 2012. These two years have significant differences in the total amount of received rainfall in the region – 2002 was a dry year and 2012 was a wet year. Using more than one growing season allows for temporal variations in the spatial relations to be examined as well. By mapping the local regression results, an effort to identify areas where Sahel is especially sensitive to variations in rainfall was made which aims to direct further research on regions with sensitive spatiotemporal dynamics. All GWR models were compared to their corresponding OLS global model. The regression analyses were performed both over the

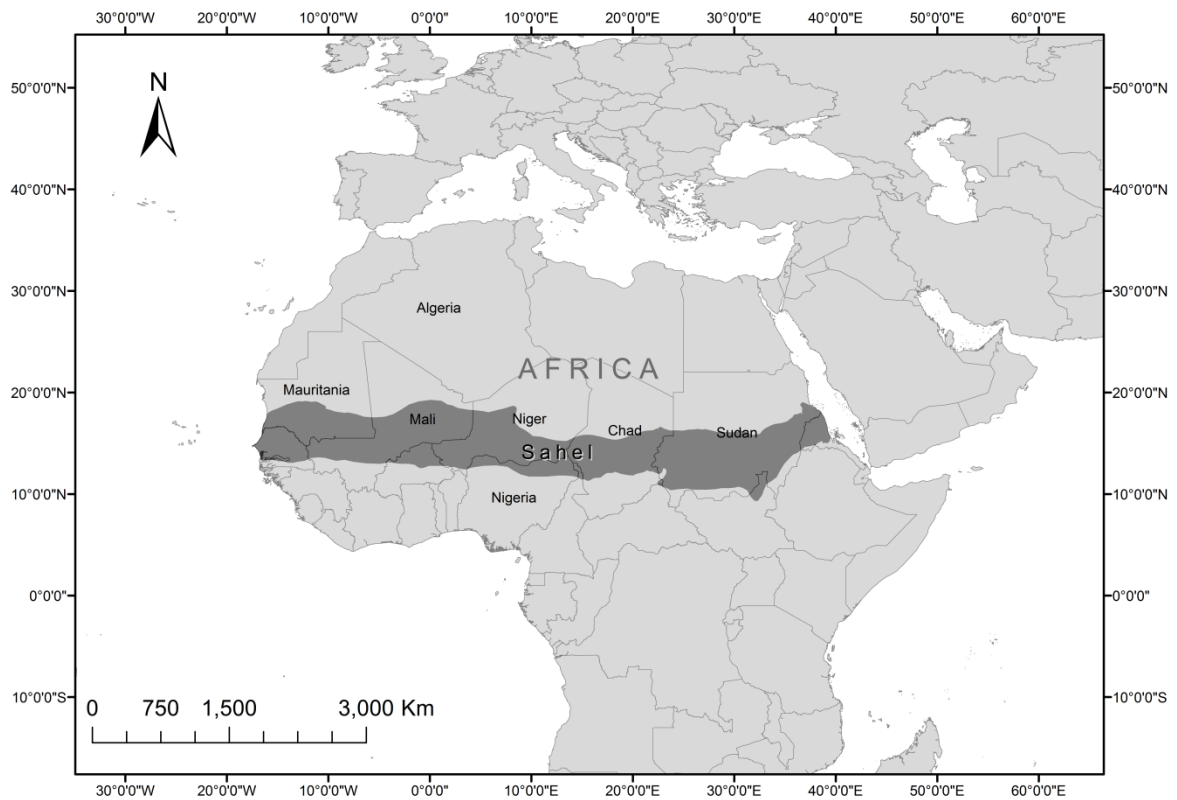
whole extent of the study area and within different land cover types. This information could be of potential use to ecologists seeking to obtain a better understanding of climate variation and vegetation distribution in the area. Another important aim is to identify the suitable scale for the local analysis. The spatial scale is expected to be influenced by the land cover distribution and vegetation type and composition in the Sahelian belt. Moreover, the GWR outputs will be used to identify areas where the model behaves poorly, providing evidence for degraded land or other factors influencing the NDVI variation such as human impact. Lastly, there are research areas which might benefit from the results of this thesis since NDVI and rainfall are often essential components in net primary production (NPP) modelling. Understanding their relationship from a more local perspective could hopefully provide fruitful information to improving the modelling of NPP itself. Similarly, increasing the prediction accuracy of estimating biomass in Sahel by undertaking local approaches to deal with potentially non-stationary data might be of benefit.

The research questions this thesis will try to address can be expressed by the following points:

- Was there evident non-stationarity in the NDVI – rainfall relationship in the Sahel and was it scale dependant?
- Did conventional regression models fail to capture the true essence of that relationship?
- Was GWR a more appropriate method to model this relationship, as assessed by standard statistical measures such as prediction accuracy and precision and residual autocorrelation while at the same time accounting for increased model complexity?
- Were there obvious temporal variations in the spatial relationship between NDVI and rainfall?
- Did GWR manage to highlight regions that emerge with extreme sensitivity in regards to the impact rainfall has in vegetation?

### 1.3 Study Area

The Sahel of Africa is an ecoclimatic transition zone between the Sahara Desert in the north and the humid Sudanian savannas of the south (Figure 1). It spans more than 5000 kilometres in length and can be as wide as 1.000 kilometres extending from the Atlantic Ocean to the Red Sea. Most of the population relies in subsistence agriculture, cash crops and livestock herding. The countries of the Sahel have a high positive population growth, which is predicted to increase from 367 million in 2000 to 1 billion by 2050 (Abdi et al. 2014).

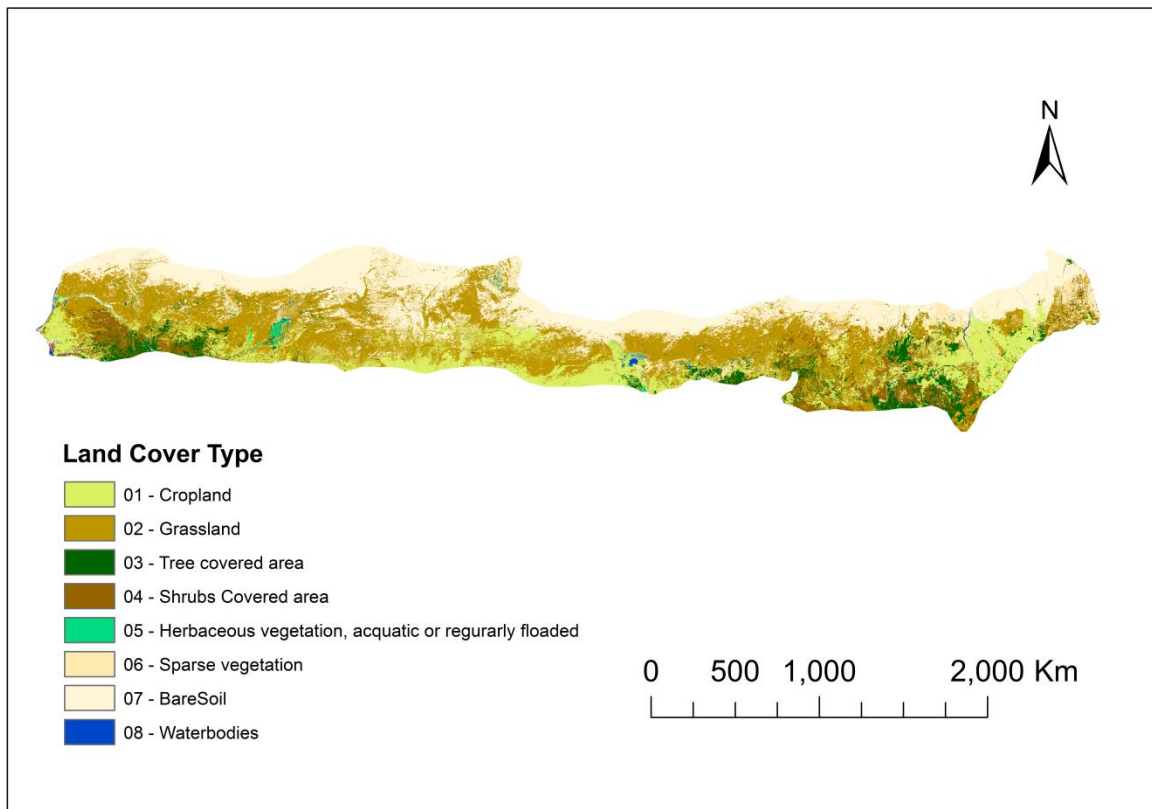


**Figure 1.** Study area. The Sahel (dark grey) is a transition zone between the arid deserts of the north to the humid savannas of the south.

The Sahel can be described as a relatively flat region covered with savannas and grasslands and comprises 11 countries. Due to its sensitivity to seasonal climatic variations its boundaries are not strictly defined. According to the Global Land Cover-

SHARE (GLC-SHARE) dataset (Figure 2) which was used in this study, the main land cover categories of the Sahelian belt are:

- Bare soil
- Croplands
- Grasslands
- Shrub covered areas
- Tree covered areas
- Sparse vegetation
- Herbaceous cover



**Figure 2.** Land cover of the Sahel based on the GLC classification (Latham et al. 2014).

The land cover pattern follows the rainfall distribution with barren areas and grasslands dominating the northern parts of the Sahel, and croplands, shrub and tree covered areas appearing more frequently in the south. The condition and amount of vegetation is



heavily controlled by rainfall which ranges from approximately 100 mm in the north to 600 - 900 mm in the south and can vary significantly each year while the temperature is generally high throughout all the year and displays very limited variation (36° - 42° C average high temperatures). The majority of the rainfall is distributed over 2-4 months during the summer growing season (usually June – September) whereas rainfall during the rest of the year rainfall is negligible (Brandt et al. 2015).

#### **1.4 Thesis Outline**

This chapter presents the rationale for this thesis, the general aims and research questions to be tackled, and some information regarding the study area.

Chapter two describes some of the literature that has focused on the vegetation-rainfall relationship in the Sahel. The latter part of the chapter focuses on discussing some of the applications of Geographically Weighted Regression in landscape ecology and will also address some of the criticisms it has received.

The first part of chapter three will outline and describe the acquainted data that were used for the analysis of this study. Then, the methodology behind Geographically Weighted Regression will be presented in detail along with the statistical tests to showcase its potentially efficiency as well as the general flow of action in the analysis of the data.

Chapter four presents the results of the analysis. The main output of the results will be in form of maps illustrating the local diagnostics for the spatial regressions that were performed. Moreover, the results of the statistical tests comparing local and global models will be presented, along with plots and graphs showcasing the predictive power of the local models.

Chapter five encloses the discussion of the results by making connections with the research questions and general aims of the thesis and encompasses additional literature to validate the findings of the study. In addition, ideas for further research on the basis of the results are discussed.

Chapter six presents the conclusions of the thesis by summing the most interesting findings and is followed by the cited literature.



## **2. Background and Literature Review**

### **2.1 Introduction**

In this section, some background information regarding the NDVI – rainfall relationship in the Sahel is provided, along with some studies that have examined the topic of spatially varying climatic relationships in various landscapes. Lastly, some of the criticism of GWR is discussed, with a focus on ecological modelling.

### **2.2 Vegetation-rainfall relationship in the Sahel**

The Sahel is a transitional zone between the Sahara desert in the north and the humid savannahs in the south (Anyamba and Tucker 2005). As a dynamic arid and semi-arid ecosystem, it is known to be sensitive to environmental factors. The Sahel was struck by devastating droughts in the 1970s and 1980s which were previously thought to be the result of past land degradation caused by human activity (e.g overgrazing, wood exploitation) (Lamprey 1988; Mensching 1990). However, as research and monitoring on the area increased, mainly attributed to the ready availability of satellite data, many studies found a significant link between reduced rainfall and vegetation condition (Anyamba and Tucker 2005; Malo and Nicholson 1990). Moreover, time series of satellite data enabled the detection of increases in vegetation greenness since the 1980s (Eklundh and Olsson 2003). Consequently, the positive vegetation trend, – the so called ”re-greening” of the Sahel is often linked to the higher amounts of rainfall since the recovery started (Hickler et al. 2005; Kaspersen et al. 2011; Olsson et al. 2005). Although, rainfall is generally thought to be one, if not the strongest, determinant for understanding the behaviour of vegetation in arid and semi-arid environments, other factors are often discussed as potentially important such as population fluctuations and migration, agricultural policy and land cover change (Fensholt and Rasmussen 2011; Olsson et al. 2005 ).

A large number of studies have explored the temporal response of vegetation condition to rainfall variability in the Sahel zone at various scales (Li et al. 2004; Olsson et al. 2005) providing rich information regarding vegetation phenology and trends. However, a smaller number of studies has focused on the spatial patterns of vegetation in response to rainfall and concluded that the NDVI-rainfall relationship emerges with different features

based on location and the variability of land cover (Chamaille-Jammes et al. 2006; Nicholson et al. 1990). These analyses are mostly based on conventional global models. In essence, even though they are computed on sampled spatial locations they are “aspatial”, since a set of regression coefficients is considered to apply equally to the whole extent of the study area (Foody 2002). The Sahel spans approximately 5,000 kilometres in width with a heterogeneous mix of land cover, vegetation composition, and soil type. Therefore, it is very likely that a global regression analysis is not able to provide realistic information about the effect of rainfall variations in the sensitivity of NDVI in the region. Even if global regression analysis are to be taken separately for different types of land cover, as Wang et al. (2001) have done, this would raise further questions. Arbitrarily defining land cover as the dominant factor of variation in the relationship might cause bias in the model, especially when the Sahelian land cover is dynamic and mixed to a degree. The quality of various land cover datasets for precise ecological analyses can also be questionable (Propastin et al. 2008). Moreover, this approach does not allow for measuring variations within the same category. Lastly, since the process of creating land cover data varies between producers, land cover maps might have significant variations between them. This further promotes the advantages of a local modelling approach, allowing a more natural way to examine the relationship between NDVI and rainfall in the Sahel.

### **2.3 Geographically Weighted Regression in ecological modelling**

The first application of the Geographically Weighted Regression between NDVI and rainfall was performed by Foody (2002). The results of his study showed statistically significant spatiotemporal non-stationarity in all regression parameters in a large area extending from Central-Northern Africa to a part of the Middle East. In addition, a significant increase in the variance explained was accomplished by selecting the GWR model over one based on an OLS regression. The findings also raised questions related the accuracy of the datasets and the calibration that is needed for a realistic approach using GWR. In his follow up, Foody (2003) illustrated the different model results produced in the GWR model by changing the spatial scale. Although his second study was related to species richness, both rainfall and NDVI data were used as predictors with varying relationships both between them and the dependant variable. Propastin et al.

(2008) and Propastin (2009) investigated the importance of finding an appropriate spatial scale in order to perform a meaningful GWR landscape analysis by employing standard statistical measures such as the AIC. This was interpreted as finding the smallest homogeneous landscape unit in the study area in order to calibrate the model in the most efficient way. In addition, a significant improvement in the decline of residual autocorrelation was accomplished by applying the GWR model to the NDVI-rainfall relationship.

Moreover, a GWR approach was shown to be superior to both OLS and SAR models in estimating net primary production (NPP) in Chinese forests by Wang et al. (2005). NPP modelling has become an important topic of research related to climate change, especially in ecological transitional zones such as the Sahel (Abdi et al. 2014) and perhaps a local modelling approach could reveal more detail about the drivers regulating NPP.

By employing GWR models in northern China, Zhao et al. (2015) showed that transitional zones from arid to humid environments are sensitive to variations in precipitation but the strength of the association varies in space even in seemingly compact land cover classes. GWR is able to transcend these issues by exploring relationships locally, by disregarding arbitrarily defined limitations. Significantly better predictions were made by the GWR approach as the global models were deemed too inefficient to properly assess the relationships under examination. As a sign of particular interest, areas where the model was miss specified were showcased and potential inferences to human activities or land degradation were made.

#### **2.4 Criticisms of Geographically Weighted Regression**

Nonetheless, GWR has received its own share of criticism, both as a methodology and an application in ecological modelling. The first critique relates mainly to issues of local multicollinearity in multivariate relationships and the extreme flexibility of the model. This has been discussed in many studies (e.g. Paez et al. 2011; Wheeler and Tiefelsdorf 2005) and is of concern when applying spatially disaggregated models. Kalogirou (2013) addressed the issue of local collinearity by developing tools to compute the correlation coefficient locally among exogenous variables. On the topic of the excessive flexibility of

the technique, essentially producing non-stationary coefficients even when using stationary datasets, Fotheringham et al. (2002) stated that insignificant variation in the regression parameters is to be expected in homogenous data. Indeed, GWR is a method that assumes stationarity in the relationships and performs statistical tests to prove otherwise. Paez et al. (2011) performed simulations on stationary controlled datasets to test this claim. In general, even though there was some variation in the parameters, it was not high. Additionally, the bandwidth of the local regression often reached a global state, meaning that with some statistical common sense, stationarity of the data can be detected by GWR. Naturally, the results were bound by the specific methods that were used. Paez et al. (2011) also signified the importance of having large datasets to receive robust regression results. As a local technique, GWR tends to be unstable when few samples are used ( $n < 140$ ), but results tend to be more stable and realistic as sample size increases.

Jetz et al. (2005) expressed concerns regarding the potential of the method to draw inferences regarding spatially varying relationships in landscape ecology and supported its use mainly as an exploratory tool. Jetz et al. (2005) questioned the potential of GWR to deal with spatially autocorrelated residuals but most studies have demonstrated the opposite (Propastin et al. 2008; Wang et al. 2005; Zhao et al. 2010). In fact, Wang et al. (2005) showed that SAR models are not nearly as effective as GWR in removing autocorrelation in the residuals. Since, relationships are established locally and directly it is only reasonable to assume that this is the case. Jetz et al. (2005) suggested that relationships are in fact global and vary locally due to missing variables or spatial interaction between variables. If this is the case, then any inferences made to explain spatially varying relationships by GWR should be taken with extreme caution. However, Wang et al. (2005) showed that relationships in their model tended to vary both in multiple bivariate and multivariate models, furtherly promoting the notion that relationships do in fact vary in space.

### **3. Data and Methodology**

#### **3.1 Data**

##### **3.1.1 Normalized Difference Vegetation Index – NDVI**

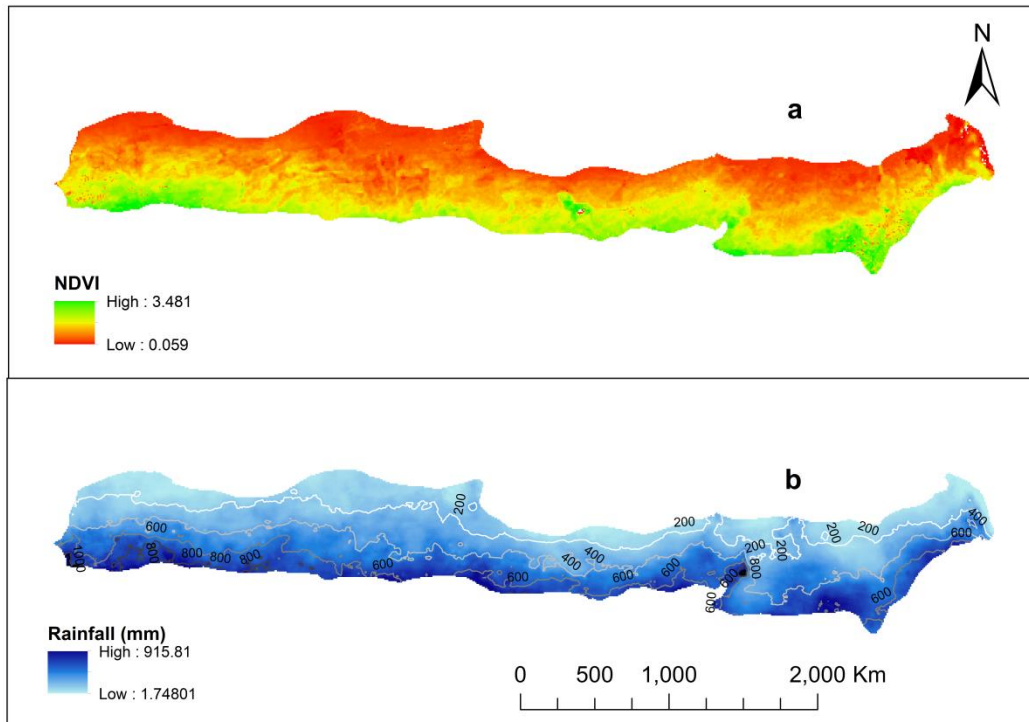
The Normalized Difference Vegetation Index (NDVI) was used as the dependant variable in this study. Satellite derived NDVI is one of the most common indices for vegetation monitoring (Tucker 1979). NDVI is computed as the ratio between two wavelengths of the electromagnetic spectrum as following:

$$NDVI = (NIR-RED)/(NIR+RED)$$

Where NIR and RED are the spectral reflectance in the near infrared (800 - 1000 nm) wavelength and in the red (620 - 750 nm) sections, respectively, of the electromagnetic spectrum. The rationale behind the use of the index is based upon the strong reflectance green vegetation displays in the near infrared portion of the spectrum while absorbing the photosynthetically active radiation (PAR) at wavelengths between in the red section. The index takes values from 0 - 1 and usually values from 0.1 up to 0.9 correspond to various stages and types of green vegetation condition. Very low NDVI values can correspond to a complete lack of vegetation, sparse desert vegetation, clouds or water bodies.

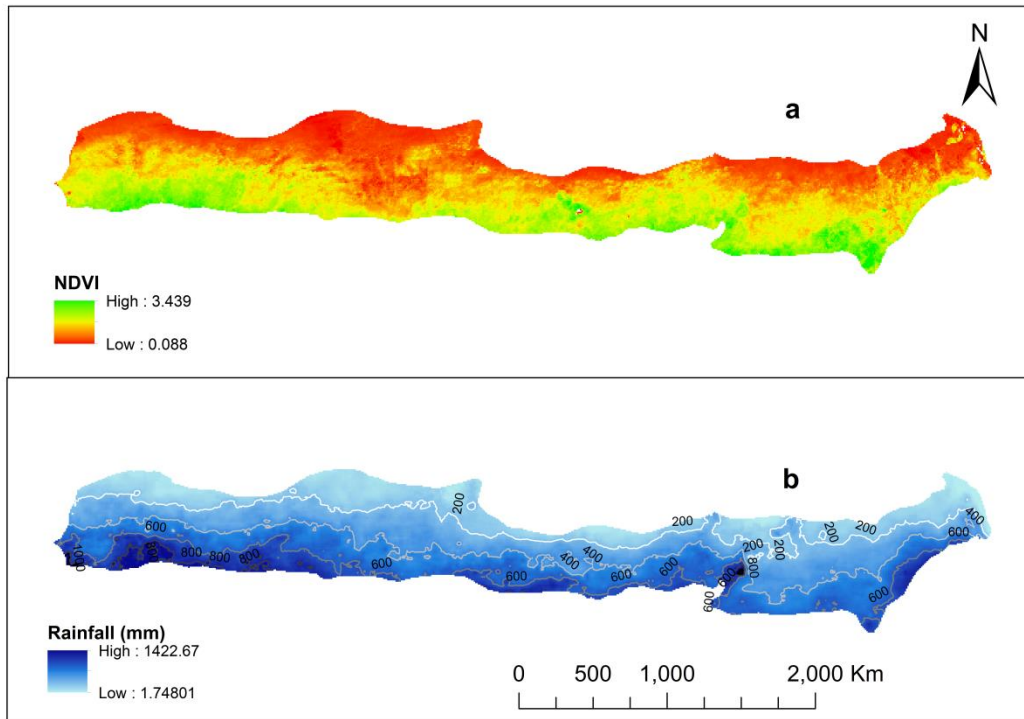
##### **3.1.2 AVHRR GIMMS NDVI**

This study utilizes the NDVI derived from the Advanced Very High Resolution Radiometer (AVHRR) instrument carried by the National Oceanic and Atmospheric Administration (NOAA) polar orbiting satellites. The dataset has been constructed, corrected and published by the Global Inventory Monitoring and Modelling System (GIMMS) project (<https://nex.nasa.gov/nex/projects/1349/>). The data are provided as biweekly NDVI images at an 8 kilometre spatial resolution with the monthly images constructed using the maximum value composite method (MVC) to reduce cloud disturbance and increase the overall quality of the dataset (Fensholt and Proud 2012). GIMMS data incorporate a plethora of corrections related to atmospheric changes, satellite sensor decay and orbit drifts (Tucker et al. 2005).



**Figure 3.** a) Integrated NDVI over June-September in 2002, b) Cumulative rainfall (mm) over June-September in 2002





**Figure 4.** a) Integrated NDVI over June-September in 2012, b) Cumulative rainfall (mm) over June-September in 2012

From these monthly maximum value composite images, growing season integrated NDVI (iNDVI) images were constructed for the growing season in the Sahel (June – September) for the years of 2002 (Figure 3) and 2012 (Figure 4). Although, the start and end of the growing season is subject to temporal variations, these four months incorporate the majority of the vegetation growth and accumulated rainfall in each year as the ratio of annual cumulative rainfall and rainfall in the growing season had a mean of 90% in 2002 and 92% in 2012.

### 3.1.3 Climate Hazard Infrared Precipitation with Stations (CHIRPS)

The Climate Hazards Group Infrared Precipitation with Station (CHIRPS) project provides rainfall datasets at the global scale since the past 30 years. The dataset is provided at a  $0.05^\circ$  degree resolution and is often used for creating rainfall time-series, monitoring droughts and associating it with vegetation data such as NDVI. Based on a conjunction of in-situ measurements and satellite data it incorporates various

interpolation techniques to produce robust precipitation grids over the globe (Funk et al. 2015).

First, the rainfall data were processed to be in accordance with the NDVI spatial resolution (8km). Then, growing season cumulative rainfall (CR) images were produced for 2002 and 2012 as the independent variables in order to estimate the relationship between NDVI and rainfall during the growing season in the Sahel.

### 3.2 Methodology

This section provides a description of the tools that were used in the analysis of this study. As ordinary least squares (OLS) regression is widely used and well known in scientific literature, the focus will be on presenting the framework for the geographically weighted regression (GWR) technique.

#### 3.2.1 Geographically weighted regression

The complete theoretical background and description of the algorithm is presented in Fotheringham et al. (2002). GWR is a geographical extension of the traditional global regression framework. Equation 1 describes a simple bivariate linear regression:

$$y_i = b_0 + b_1 x_i + \varepsilon_i, \quad i = 1:n \quad (1)$$

where  $y$  is the dependant or response variable;  $x$  the independent or repressor variable;  $\varepsilon$  the error term;  $b_0$  and  $b_1$  are the alpha (intercept) and beta (slope) coefficients, respectively; and  $n$  is the number of samples that correspond to spatial locations. In the OLS framework  $b_0$  and  $b_1$  are estimated in such a way that the sum of square residuals based on the spatial samples would be minimized (Brunsdon et al. 1998). Moreover, the residuals should satisfy both spatial independence (the observations should be independent of each other) and homoscedasticity criteria (the variability of the dependent variable is constant across ranges in respect to the predictor variable). The estimation of these parameters can be described as a set of matrix equations as in Equation 2:

$$\hat{b} = (X^T X)^{-1} X^T y \quad (2)$$

where  $\hat{b}$  is the vector of estimated parameters;  $X$  is the matrix calibration that contains independent variables;  $y$  is the vector of dependent variables and  $(X^T X)^{-1}$  equals the inverse of the variance – covariance matrix (Fotheringham et al. 2002).

The regression parameters are constant and are assumed to apply to any location from which the samples were taken. The coefficient  $\hat{b}$  indicates the rate of change in the dependent variable with a unit change in the independent variable. In the global framework this rate is universal at all spatial locations. In this study, this can be described as the rate of change in NDVI with an increase or decrease in rainfall. GWR extends Equation 1 in order to incorporate locational information where  $b_1(u_i, v_i)$  is the  $i$ th parameter for location  $u, v$  (describing positional information such as latitude and longitude) (Brunsdon 1998), as shown in Equation 3.

$$y_i = b_0(u_i, v_i) + b_1(u_i, v_i)x_i + \varepsilon_i, \quad i = 1:n \quad (3)$$

In this case, the parameters are computed as a function of location and are allowed to vary through space. Essentially, a different estimate is produced for each spatial observation. The calibration of these parameters is based by performing a weighted least squares (WLS) regression on nearby observations for every  $u, v$ . Consequently Equation 2, can be rewritten as:

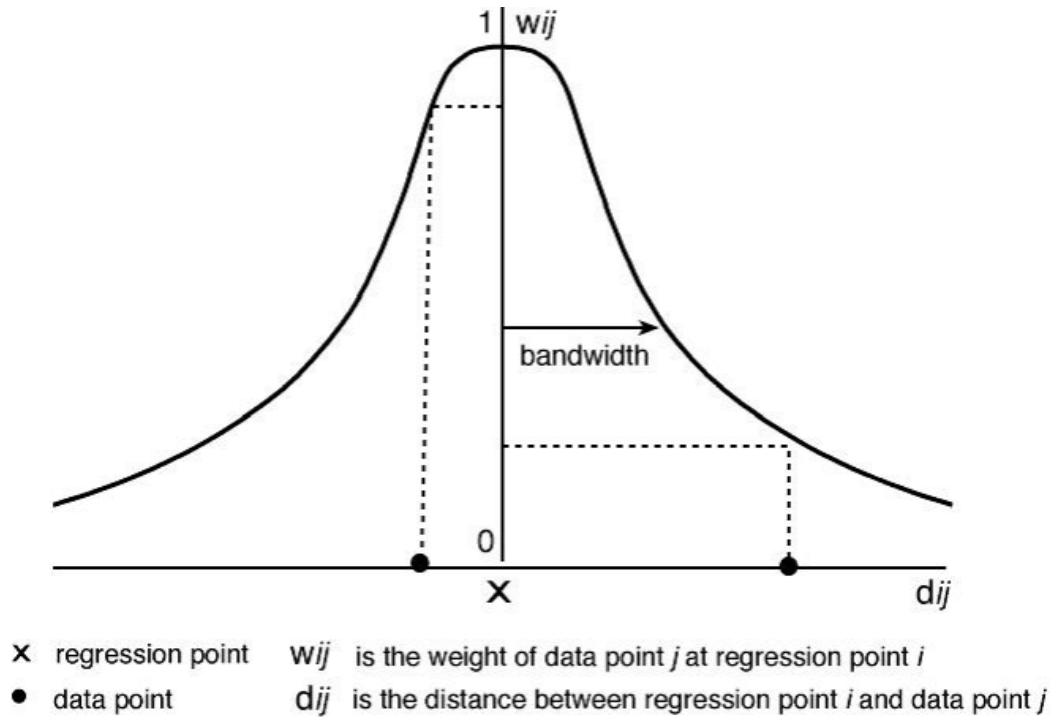
$$b_1(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y \quad (4)$$

where  $W(u_i, v_i)$  is a weight matrix that assigns different weights to observations near each regression point  $i$ . Observations closer to the location of each calibration are assigned higher weights, while those further away are assigned lower weights. This gives an explicit spatial character to the local regressions since observations near  $i$  take on a heavier burden in the prediction accuracy (Fotheringham et al. 2002). Therefore, spatial variations in the relationship between dependant and exogenous variables can be mapped to reveal potentially interesting local variations.

The type of spatial weighting is important in a GWR calibration. The most commonly used types are Gaussian, or Gaussian-like, forms of spatial weighting:

$$W(i, j) = \exp -\left(\frac{d_{ij}}{b^2}\right) \quad (5)$$

where  $W(i, j)$  is the weight of observation  $j$  for location  $i$ ;  $d$  denotes the Euclidian distance between these points and  $b$  is the size of the kernel (bandwidth). This allows for estimating parameters even at locations where we do not have sample data precisely because GWR computes parameters as a function of weighting nearby observations (Figure 5). For example, if there exists an observation  $j$  at regression point  $i$  the weight will be one while if it is further away it will follow the Gaussian distance decay function described by Equation 5. The bandwidth size ( $b$ ) describes the distance limit in which nearby observations are taken into account for calibrating the regression. If  $b$  is too large, the GWR model would turn into the global model since all observations would be used at each local regression, thus hindering meaningful parameter variation and increasing model bias. On the contrary, if  $b$  is too small the parameters will be less biased but will depend on few nearby samples and will have increased variance in their estimates and large standard errors due to low degrees of freedom (Propastin 2009). Therefore, the results of GWR become increasingly dependent on the choice of  $b$ , deeming it perhaps the most important calibration value. Essentially, the choice of the bandwidth is a choice between goodness of fit and degrees of freedom (Karen 2008) and between variance and bias (Figure 6). The bandwidth can either be of a *fixed* or *adaptive* nature. A fixed bandwidth describes a kernel of a predetermined distance size that continuously moves invariantly through the study area. The number of observations is dependant to their distribution as in some cases sample density might be too high or too low. In the *adaptive* scheme the bandwidth performs as a function of predetermined number of nearest neighbours and can vary in size in order to incorporate the same number of observations in each iteration. Both weighting schemes, and Gaussian weighting, were used in this study.



**Figure 5.** Gaussian Spatial Weighting function for GWR (Propastin et al. 2008)

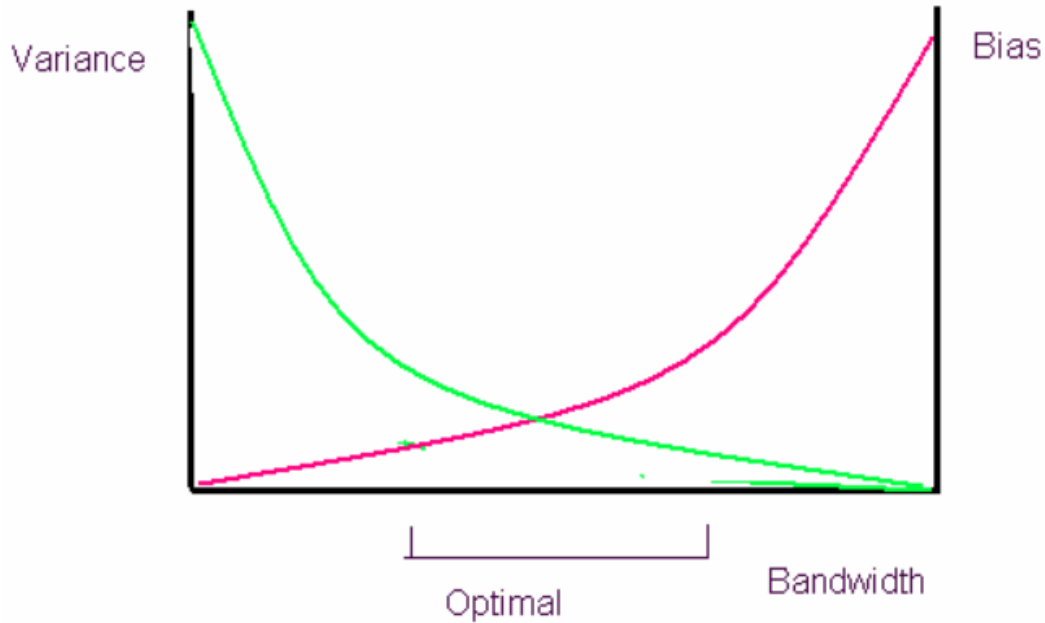
### 3.2.2 Scale dependency of non-stationarity

In landscape ecology, in order to obtain reliable results from GWR and select a meaningful bandwidth size, the *Stationarity Index (SI)* proposed by Osborne et al. (2007) is often used to study the scale dependency of nonstationarity between ecological variables. The SI is calculated as:

$$SI = \frac{iqr\_gwrse}{SE*2} \quad (6)$$

where,  $SI$  is the stationarity index,  $iqr\_gwrse$  is the interquartile range of standard errors for GWR coefficients and  $SE$  is the standard error for a coefficient in a global model such as OLS. Values less than 1 indicate stationarity at that scale, implying that the relationship has stabilized. The SI can illustrate the difference between the observation and intrinsic scale between NDVI and rainfall in the Sahel. The observation scale describes the whole study area in which spatial constancy is assumed, whereas the intrinsic scale relates to the operational scale in the ecological relationship, which can be

of a much smaller size (Gao and Li 2011). In this study, the SI was calculated at multiple scales ranging from 25 km to 900 km in order to study the variation of scale dependency in the local parameters.



**Figure 6.** Selection of bandwidth size (Fotheringham et al. 2002)

### 3.2.3 Model comparison

In order to validate the use of GWR over a global model and over different bandwidths statistical tests were used to test its potential improvement. The Akaike Information Criterion corrected for small sample sizes (AICc) is a standard statistical measure to compare the relative performance of two models by accounting for model complexity and differences in degrees of freedom (Akaike 1974):

$$AICc = 2n \ln(\hat{\sigma}^2) + n \ln(2\pi) + n \left( \frac{n - \text{tr}(S)}{n - 2 - \text{tr}(S)} \right) \quad (7)$$

where  $n$  is the sample size,  $\sigma$  is the deviation of the residuals and  $\text{tr}(S)$  is the trace of the hat matrix, the matrix that transforms observations into predictions in regression terms

(Karen 2008). In general, a decrease in AICc of more than 3 indicates a better model. Moreover, an F-test (Equation 8) proposed by Fotheringham et al. (2002), that is based in an analysis of variance (ANOVA) was computed to assess the significance of the improvement:

$$F = \frac{RSS_{gwr}/DF_{gwr}}{RSS_{glm}/DF_{glm}} \quad (8)$$

where,  $RSS_{gwr}$  is the residual sum of squares for a GWR model;  $RSS_{glm}$  is the residual sum of squares for a global model;  $DF_{gwr}$  and  $DF_{glm}$  are the degrees of freedom for GWR and the global model, respectively. Finally, the Root Mean Square Error (RMSE) was computed for the global and local models, respectively.

### 3.2.4 Tests for the significance of spatial non-stationarity

In order to assess if the regression parameters show significant spatial variation the geographical variability test (Nakaya et al. 2009) was applied. The test examines the spatial variability of parameters by comparing a traditional GWR model with a model in which every  $k$ th coefficient is kept constant while the rest of the coefficients are allowed to vary e.g.:

$$GWR \text{ model: } y_i = b_o(u_i, v_i) + b_1(u_i, v_i)x_i + \varepsilon_i \quad i = 1:n$$

$$Fixed \text{ model } 1: y_i = b_o(u_i, v_i) + b_1x_i + \varepsilon_i, \quad i = 1:n$$

$$Fixed \text{ model } 2: y_i = b_o + b_1(u_i, v_i)x_i + \varepsilon_i, \quad i = 1:n$$

where *GWR model* is that model being tested, *Fixed model 1* is the test for the slope coefficient and *Fixed model 2* is the test for the intercept coefficient. The indicator for the significance of the spatial variation in the regression parameters is the AICc, used by comparing the values for different models. Moreover, a more exploratory method was presented by comparing the interquartile range of GWR coefficients that account for 50% of the data distribution with the global estimate  $\pm$  one standard deviation which accounts for about 68% of the data distribution since in a global model the parameter estimate is

the mean of a Gaussian curve (Fotheringham et al. 2002). If the local estimates exceed that limit we can conclude important spatial variation in the coefficients.

### 3.2.5 Testing for residual autocorrelation

A popular statistic to assess the degree of spatial autocorrelation of a data set is the Moran's I index (Moran 1948). Moran's I measures the degree of association between a vector of observed values and the weighted average of their neighbours. In this thesis the formula proposed by Cliff and Ord (1973) was used:

$$I = \frac{n \sum_i^n \sum_j^n w_{ij} z_i z_j}{M \sum_{i=1}^n z_i^2} \quad (9)$$

where  $n$  is the number of data points,  $z_i = x_i - \bar{x}$ ,  $\bar{x}$  is the mean value of  $x$ ,  $M = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$  and  $w_{ij}$  is the element of the matrix of spatial proximity  $M$ , which depicts the degree of spatial association between the points  $i$  and  $j$  (Kalogirou and Hatzihristos 2007). In the above description  $x$  can describe the value of the variable we are testing for spatial autocorrelation – in this case the residuals. The matrix of spatial proximity was constructed by using the  $k$  nearest neighbours approach. In detail,  $w_{ij} = 1$  when  $j$  is one of the  $k$  nearest neighbours of  $i$  and  $w_{ij} = 0$  elsewhere (Anselin 2003). Various weight matrices were constructed including different number of nearest neighbours each time based on difference distance lags. More specifically, spatial correlograms were constructed, that computed the value of Moran's I index in spatial lags from 25-500km using 15 km intervals. The indices were tested against the randomization hypothesis at a 95% confidence level by using spatial permutations.

### 3.2.6 Data processing and software

After removing missing data values and NDVI values  $< 0.1$  to reduce unwanted signal coming for potentially non vegetated pixels (Jamali et al. 2014) the NDVI and precipitation raster surfaces were converted to vector format in order to use the GWR tool in ArcGIS 10.3, the lctools (Kalogirou 2015) package in the open source statistical software R 3.1.2 and the GWR4 software (Nakaya et al. 2014). Then, random sampling was performed to encapsulate 30% of the dataset corresponding to approximately 10,000 points for each year (300 less pixels in 2012).



### **3.2.7 Land cover analysis**

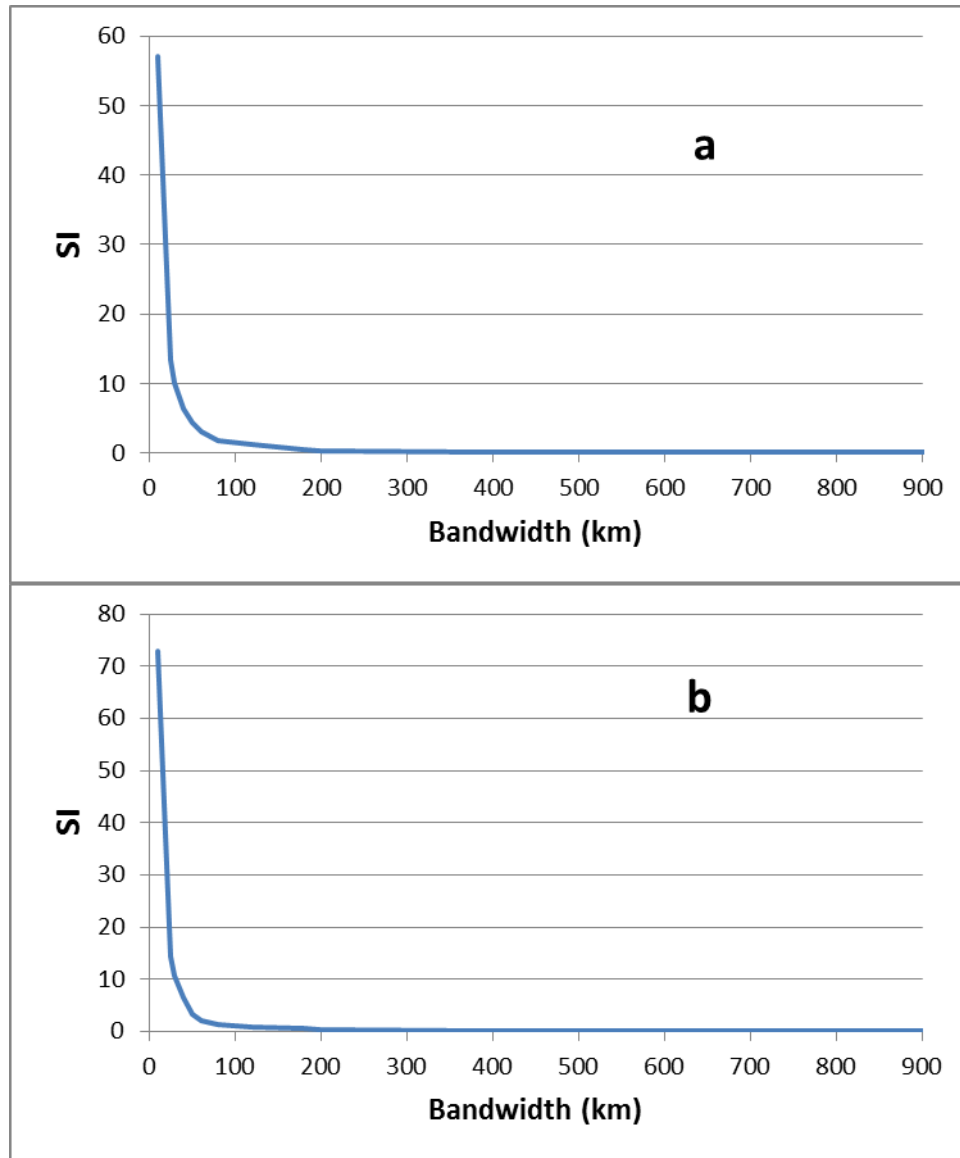
In addition to applying the regression models to the whole extent of the Sahel, both GWR and OLS models were applied within the main land cover categories based on the GLC-SHARE land cover data set (Figure 2) in order to assess the effect of land cover in the variation of the NDVI-rainfall relationship. Pixels classified as herbaceous cover were not considered due to their very small number in comparison with the rest. The GWR models employed the *adaptive* kernel type due to land cover not being evenly distributed across the study area.



## **4. Results**

### **4.1 Scale dependency**

The relationship between NDVI and accumulated rainfall for the growing season was scale dependent for the Sahel. By fluctuating the size of the bandwidth varying results were produced. The patterns were more homogenized as the bandwidth broadened and incorporated information from locations far away, thereby smoothing the regression coefficients and bringing them closer to those of a global model. On the contrary, with smaller bandwidths very detailed patterns were produced at the cost of increased standard errors. The stationarity index (SI) for both 2002 and 2012 (Figure 7) suggests that different levels of scale dependent non-stationarity were present by varying the scale of examination. However, the SI had temporal differences as well. In 2012, for small bandwidths the stationarity index was higher than the values for 2002 but became stationary slightly earlier than in 2002. The SI slopes declined abruptly with an increase in the bandwidth that stabilized around 180 kilometers implying that this is the intrinsic scale of the NDVI-rainfall relationship, i.e. the minimum geographical area which we can build a stationary and reliable relationship for the entire study area. This could be interpreted as the size of a landscape unit that describes the natural arrangement of the phenomenon where we can incorporate variations of non-stationarity while also removing unnecessary bias and noise in the model (Brunsdon et al. 1998). Consequently, as the SI slopes became flat at approximately 200 km, this was selected as the appropriate bandwidth to model the NDVI-rainfall relationship in the Sahel during the growing season.

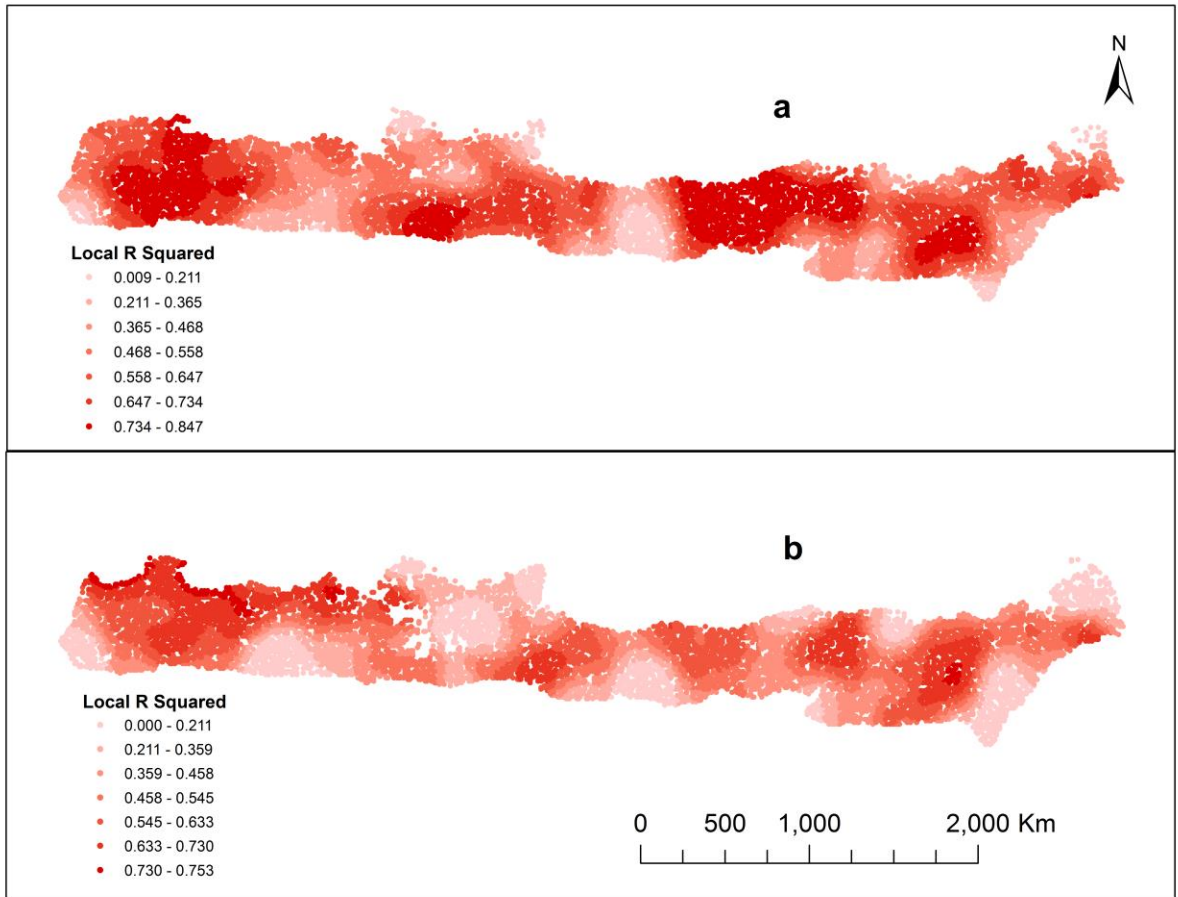


**Figure 7.** Stationarity index (SI) for 2002 (a) and 2012 (b) at multiple bandwidths. The SI is that ratio of the interquartile range of standard errors for the GWR coefficients with twice the standard error of a spatially constant, global model (GLM). Values less than 1 on the y-axis indicate stationarity at the corresponding spatial scale.

#### 4.2 Spatial patterns of the NDVI – rainfall relationship

As the regression parameters were allowed to vary through space, different temporal and spatial patterns regarding the strength of the correlation and the significance of the relationship were produced. By mapping the local diagnostics and coefficients the spatial heterogeneity and nonstationary relationship between NDVI and rainfall were illustrated.

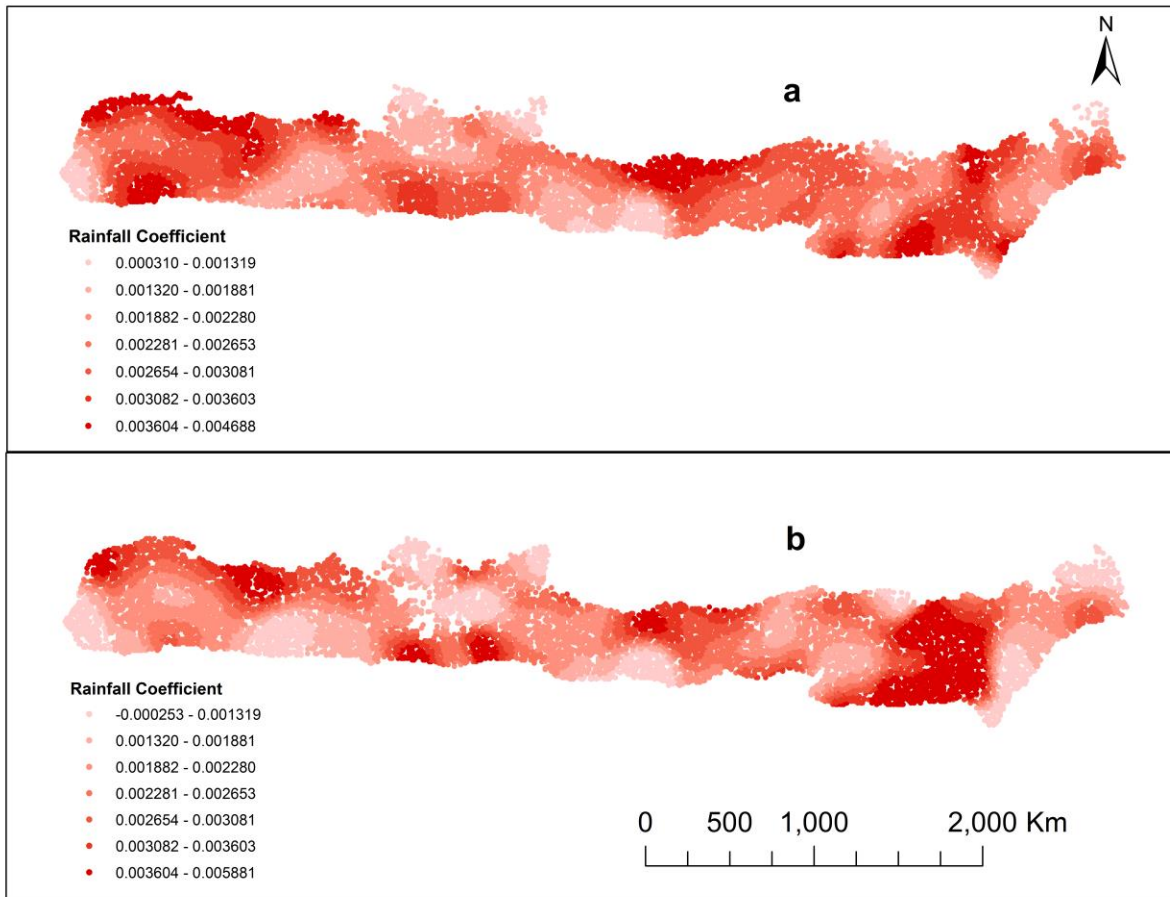
Local  $R^2$  values describe how well the local models fit the observations and indicate whether or not the model performs well. There were significant temporal and spatial variations in the strength of the correlation throughout the Sahel (Figure 8). The local fits were lower in the humid wetlands around Lake Chad and Niger River in Mopti, Mali in the central and southwest parts of the Sahel respectively. The influence of rainfall in the variations of NDVI in the growing season is low and perhaps other ecological factors have stronger impact in these areas. High correlations ( $R^2 > 0.5$ ) were found in western Sahel (parts of Mauritania and Senegal) and eastern Sahel (parts of Chad and most of Sudan) suggesting that rainfall is a very potent determinant in these areas. The local fits were higher in 2002, which was a drier year than the relatively wet 2012. The abundance of rainfall makes the correlations weaker, although the general patterns of high and low clustering remain similar in both years.



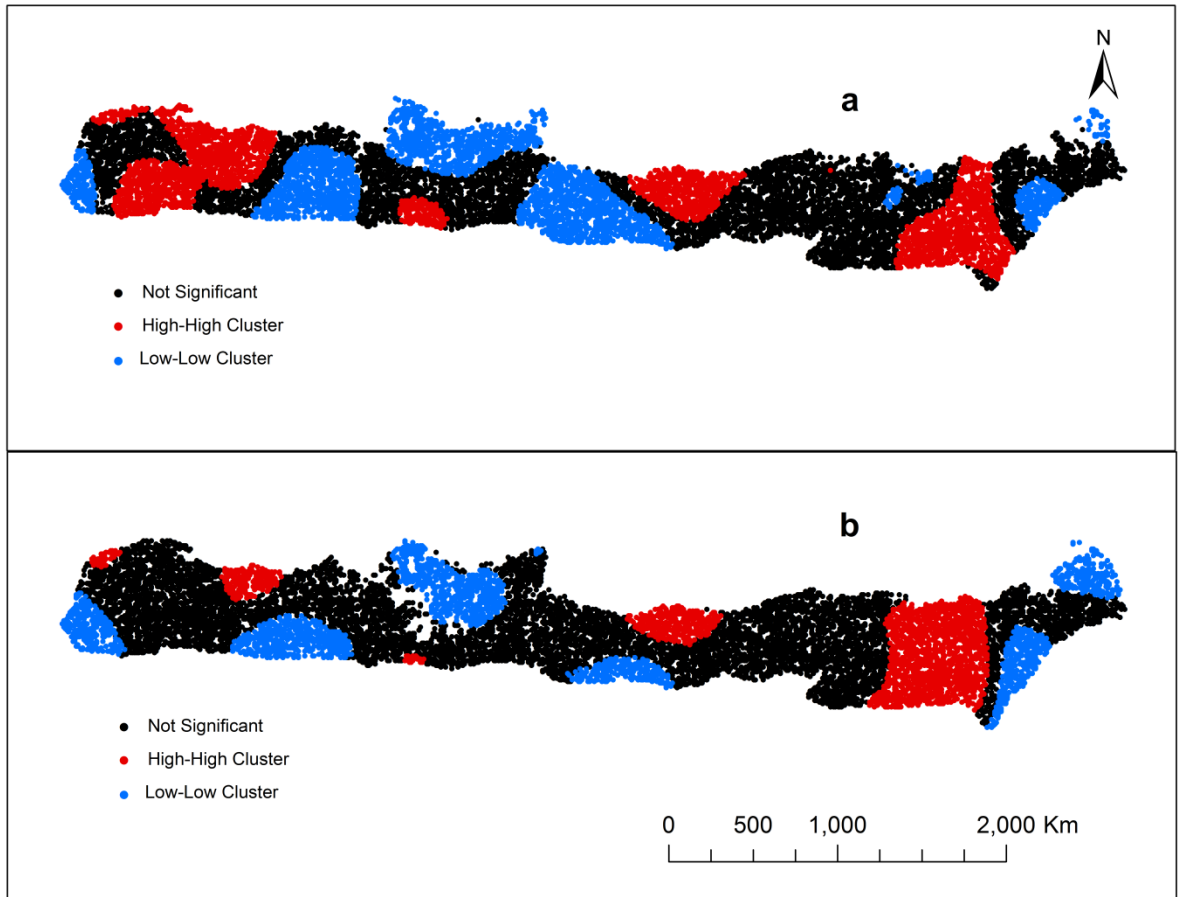
**Figure 8.** Local  $R^2$  patterns in 2002 (a) and 2012 (b). Low values indicate poor model performance and an indication for potentially missing variables in the model.

The slopes of the local regressions (beta coefficients) describe the magnitude and nature of the relationship. The vast majority of the coefficients are positive suggesting that an increase in rainfall relates to an increase in NDVI. However, the rate of increase differs significantly throughout the Sahelian belt (Figure 9). The strength of the associations was higher in the growing season of 2002 than of 2012. Nonetheless, some very important variations appear in the maps. Although the rainfall coefficient was lower in most of the Sahel in 2012 than during drier year of 2002, a seemingly large cluster covering and extending over Sudan has surprisingly high values that surpass even all the coefficient values for 2002 implying that the NDVI patterns were increasingly sensitive to variations in rainfall. These interesting local variations would be lost in a global model. It seems that regardless of rainfall amount, instead of forming a sensitive geographical

transitional zone, the Sahel forms clusters that operate as transitional passages from humid to arid environments with their size being dependent on the amount of rain received. A cluster analysis as shown in Figure 10, clearly illustrates where high and low value clusters are formed.



**Figure 9.** Rainfall (Beta) coefficient in the Sahel for 2002(a) and 2012 (b). The coefficients depict the rate of change for a spatial unit of NDVI with an increase in a spatial unit of rainfall.



**Figure 10.** Spatial Clustering Maps for the slope coefficient in the Sahel in 2002 (a) and 2012 (b), respectively. Red dots suggest significant clustering of similarly high values while blue dots suggest significant clustering of low values.

### 4.3 Land cover analysis

In this section, the results from running global and local regressions in each land cover category separately are presented (Table 1). Tree covered areas and shrublands demonstrated the highest fits while barrens and croplands signified low correlations. The local models amplified the variations within the same land cover and strengthened the relationships. This is further demonstrated by the drop in AICc values in all cases. However, the results were not surprising considering that land cover categories are mixed and span across vast distances in the Sahel. This further promotes a geographically-weighted approach for modelling NDVI and rainfall in the Sahel that takes into account



local climate and sub-regional variations that are not restricted by arbitrary limitation of land cover.

**Table 1.** Regression Correlation Coefficients and AICc for various land cover categories in the Sahel

	$R^2_{GLM}$	$AICc_{GLM}$	Min $R^2_{GWR}$	Mean $R^2_{GWR}$	Max $R^2_{GWR}$	$AICc_{GWR}$
<b>Land Cover Type</b>						
Barrens <sub>2002</sub>	0.356	-1662.16	0.167	0.448	0.707	-2709.96
Barrens <sub>2012</sub>	0.257	820.80	0.046	0.327	0.653	198.66
Shrub Covered Areas <sub>2002</sub>	0.619	477.51	0.133	0.590	0.858	34.679
Shrub Covered Areas <sub>2012</sub>	0.480	740.37	0.128	0.463	0.793	275.47
Tree Covered Areas <sub>2002</sub>	0.706	284.43	0.144	0.624	0.845	-22.12
Tree Covered Areas <sub>2012</sub>	0.526	597.63	0.173	0.478	0.764	276.67
Croplands <sub>2002</sub>	0.358	1314.02	0.000	0.416	0.759	-27.01
Croplands <sub>2012</sub>	0.312	1821.96	0.000	0.325	0.694	679.42
Grasslands <sub>2002</sub>	0.564	731.42	0.303	0.642	0.776	-1747.62
Grasslands <sub>2012</sub>	0.472	2786.35	0.217	0.508	0.652	991.21
Sparsely vegetated <sub>2002</sub>	0.732	29.04	0.184	0.545	0.874	-520.14
Sparsely vegetated <sub>2012</sub>	0.614	523.64	0.012	0.355	0.809	15.65

#### 4.4 Model Diagnostics and tests for non-stationarity

The results of the global regression for the whole study area are presented in Table 2. Both the intercept (alpha coefficient) and slope (beta coefficient) were statistically significant in regards to their relationship with NDVI ( $p < 0.01$ ). In 2002, about 69% of the variance was explained by precipitation patterns in the Sahel while 62% was explained in 2012. Global regression diagnostics implied problems with non-stationarity and residual heteroscedasticity (Koenker Statistic,  $p < 0.01$ ) and indicated potential model miss specification. On the contrary, GWR explained more of the variance (80% in 2012 and 85% in 2002) and lowered the AICc values, which accounts for changes in model complexity and degrees of freedom. The F-test based on an ANOVA suggested that GWR produced statistically significant improvement over the OLS models in both years ( $p < 0.01$ ).

**Table 2.** Model comparison between GWR and OLS models in 2002 and 2012.

<b>Model</b>	<b>R<sup>2</sup></b>	<b>AICc</b>	<b>F-test</b>
<b>GLM 2002</b>	0.688	4416.60	
<b>GLM 2012</b>	0.622	9545.55	
<b>GWR 2002</b>	0.858	-3812.01	240( $p < 0.01$ )
<b>GWR 2012</b>	0.808	2403.34	219( $p < 0.01$ )

The geographical variability test implied that both the intercept and rainfall coefficients were significantly varying in both years. This can be more illustrative by examining the interquartile range of the coefficients with the global estimates and their standard deviations.

**Table 3.** Descriptive statistics for regression parameters in 2002.

<b>Model</b>	<b>Statistics</b>	<b>Intercept</b>	<b>Slope</b>
<b>GLM</b>	Estimate	0.328622	0.002848
	Standard Error	0.00593	0.000018
	Estimate + 1 SD	0.334552	0.002866
	Estimate - 1 SD	0.322692	0.00283
<b>GWR</b>	Mean	0.362463	0.002731
	Minimum	-0.02677	0.000310
	Lower Quartile	0.256978	0.002452
	Median	0.351192	0.002789
	Upper Quartile	0.454996	0.003047
	Max	0.900006	0.004688

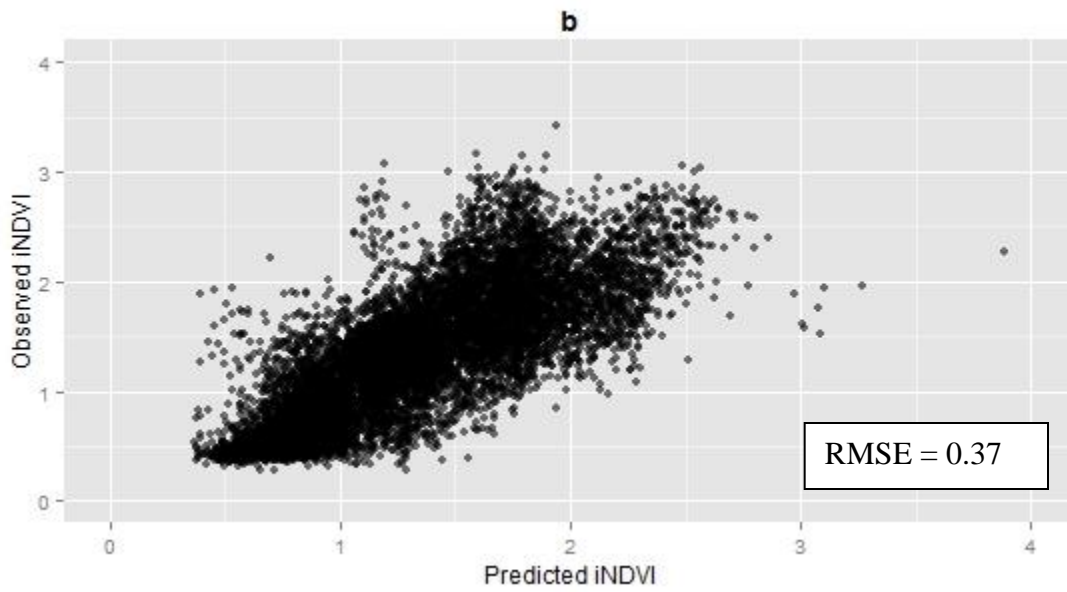
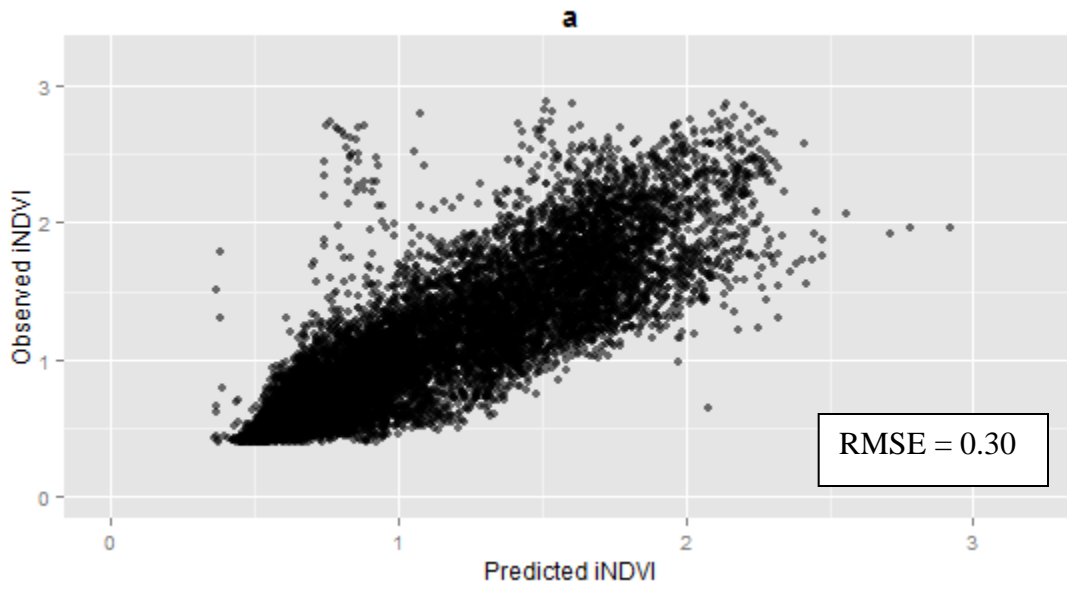
The interquartile range for GWR parameters corresponds to 50% of the data distribution while 68% of the data is found within one standard deviation for the GLM model. It is clear that in both growing seasons, regression parameters “drift” far more than the  $\pm 1$  SD of the global estimates. In 2002 (Table 3), the intercept for the global model is estimated between 0.322692 - 0.334552 while for GWR the interquartile range is estimated between 0.256978 - 0.454996. Similarly, the slope parameter for the global model is found within 0.00283 - 0.002866 while for GWR the range is 0.002452 - 0.003047. The results are similar for 2012 (Table 4) and strongly suggest that the global model captures just an average impression of the relationship in the Sahel, hindering detection of local variations and potentially supporting misleading assumptions and interpretation.

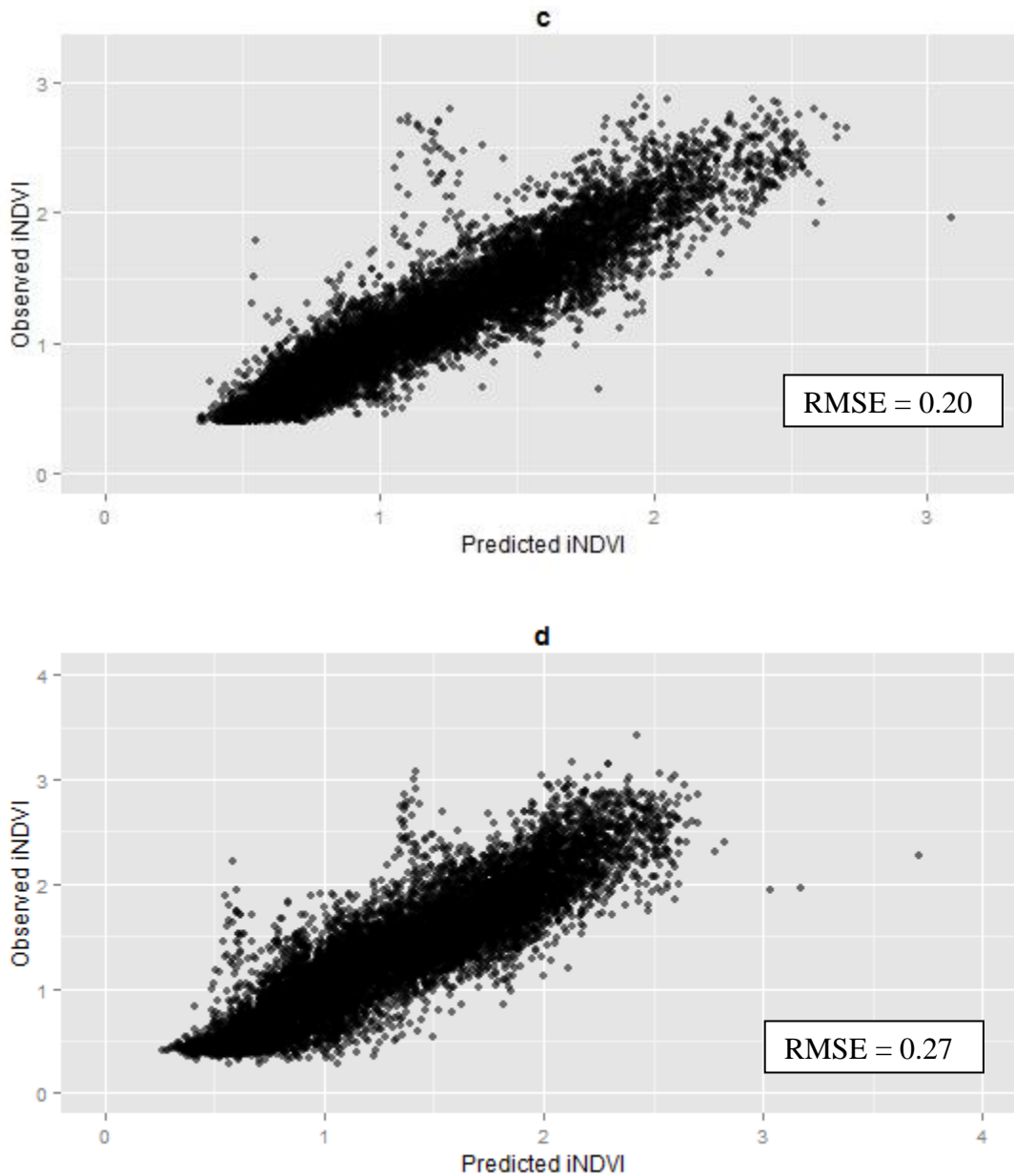
**Table 4.** Descriptive statistics for regression parameters in 2012

<b>Model</b>	<b>Statistics</b>	<b>Intercept</b>	<b>Slope</b>
<b>GLM</b>	Estimate	0.328645	0.002514
	Standard Error	0.008088	0.000019
	Estimate + 1 SD	0.336728	0.002533
	Estimate - 1 SD	0.320552	0.002495
<b>GWR</b>	Mean	0.349138	0.002472
	Minimum	-0.24613	-0.000253
	Lower Quartile	0.223988	0.002158
	Median	0.369608	0.002368
	Upper Quartile	0.489719	0.002612
	Max	0.934338	0.005881

#### **4.5 Predicted Patterns**

The scatterplots of the observed and simulated values are presented in Figure 11. The global models were not consistent and would underestimate high values of NDVI and overestimate low values. This is to be expected since a constant set of parameters was used to capture a relationship in a very large area with high spatial heterogeneity. The local modelling approach managed to produce more accurate estimates by taking into account local characteristics and incorporating locational information.

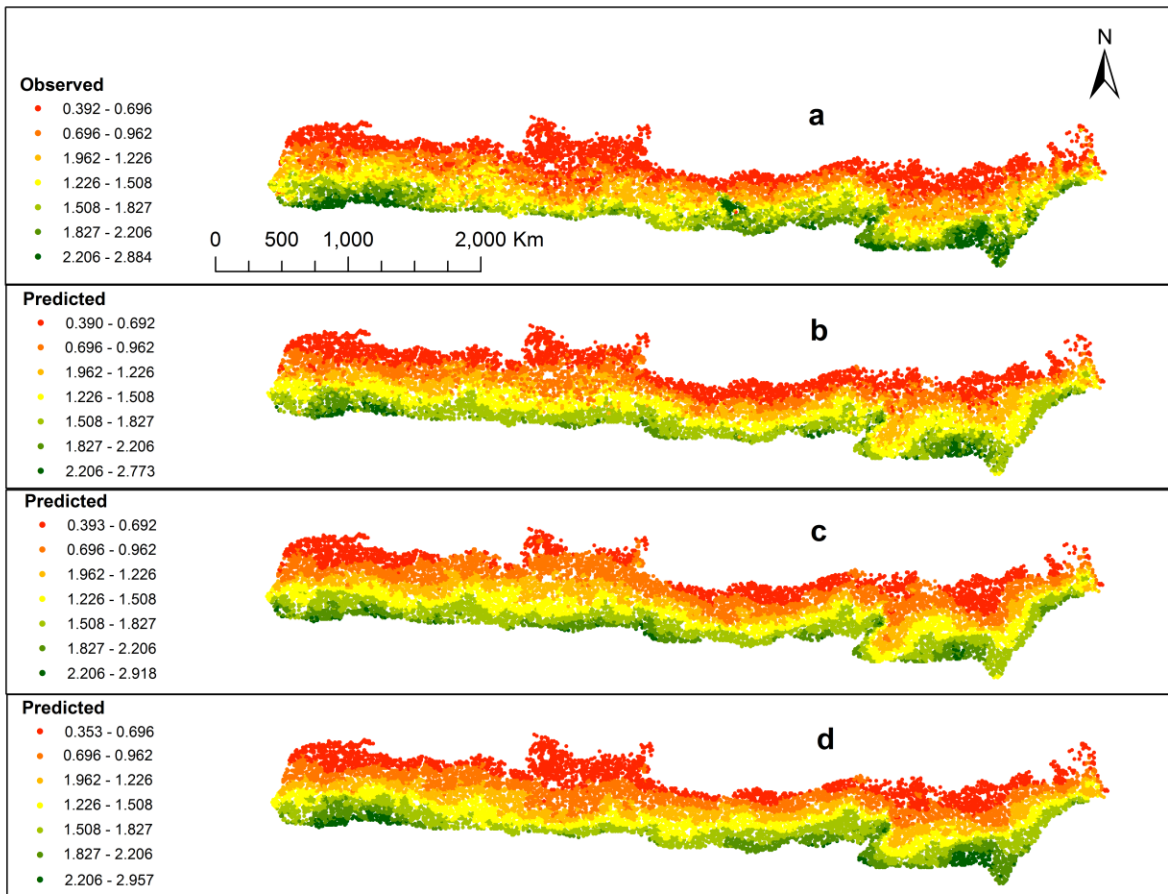




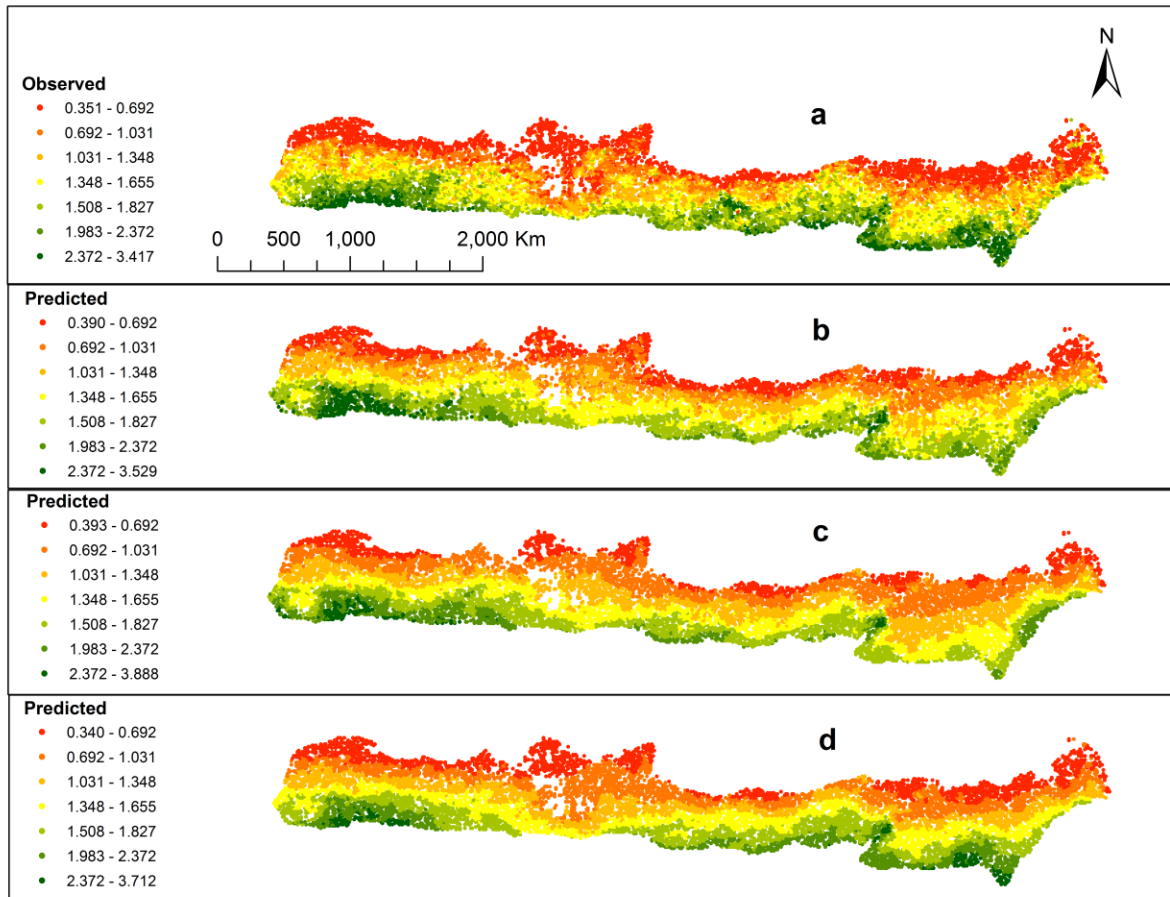
**Figure 11.** Scatter plots of observed and simulated NDVI in both growing seasons. a) OLS model in 2002, b) OLS model in 2012, c) GWR model in 2002, d) GWR model in 2012. RMSE refers to the Root Mean Square Error.

Figures 12 and 13 depict the observed and predicted spatial patterns of NDVI for 2002 and 2012, respectively. The OLS model that included data from the whole study area produced more generalized patterns that hindered any local variation in the values of

NDVI across the Sahel (RMSE =0.30 in 2002, RMSE = 0.37 in 2012). Similarly, the predictions produced by combining OLS models in different land covers produced only slightly better results (RMSE= 0.27 in 2002, RMSE = 0.35 in 2012). Both models presented the actual gradient of lower iNDVI values in the north and higher values to the south. However, the GWR model predicted it in a more precise manner because it takes into account the spatial variation in the NDVI – rainfall relationship and regional information (RMSE=0.20 in 2002, RMSE = 0.27 in 2012).



**Figure 11.** Observed and Predicted spatial patterns of iNDVI in 2002. a) Observed patterns, b) OLS models in separate land cover categories, c) Global OLS model, d) GWR model.



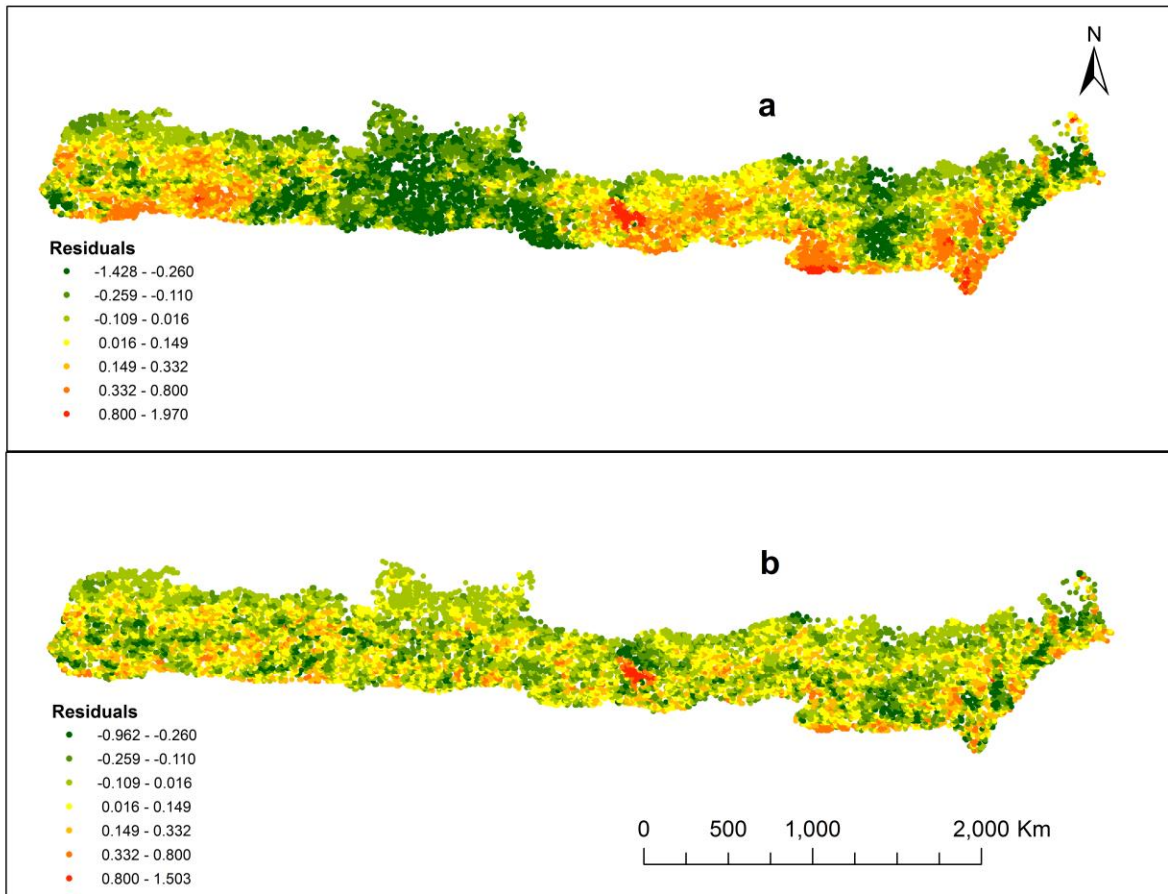
**Figure 12.** . Observed and Predicted spatial patterns of NDVI in 2012. a) Observed patterns, b) OLS models in separate land cover categories, c) Global OLS model, d) GWR model.

#### 4.6 Spatial autocorrelation of the residuals

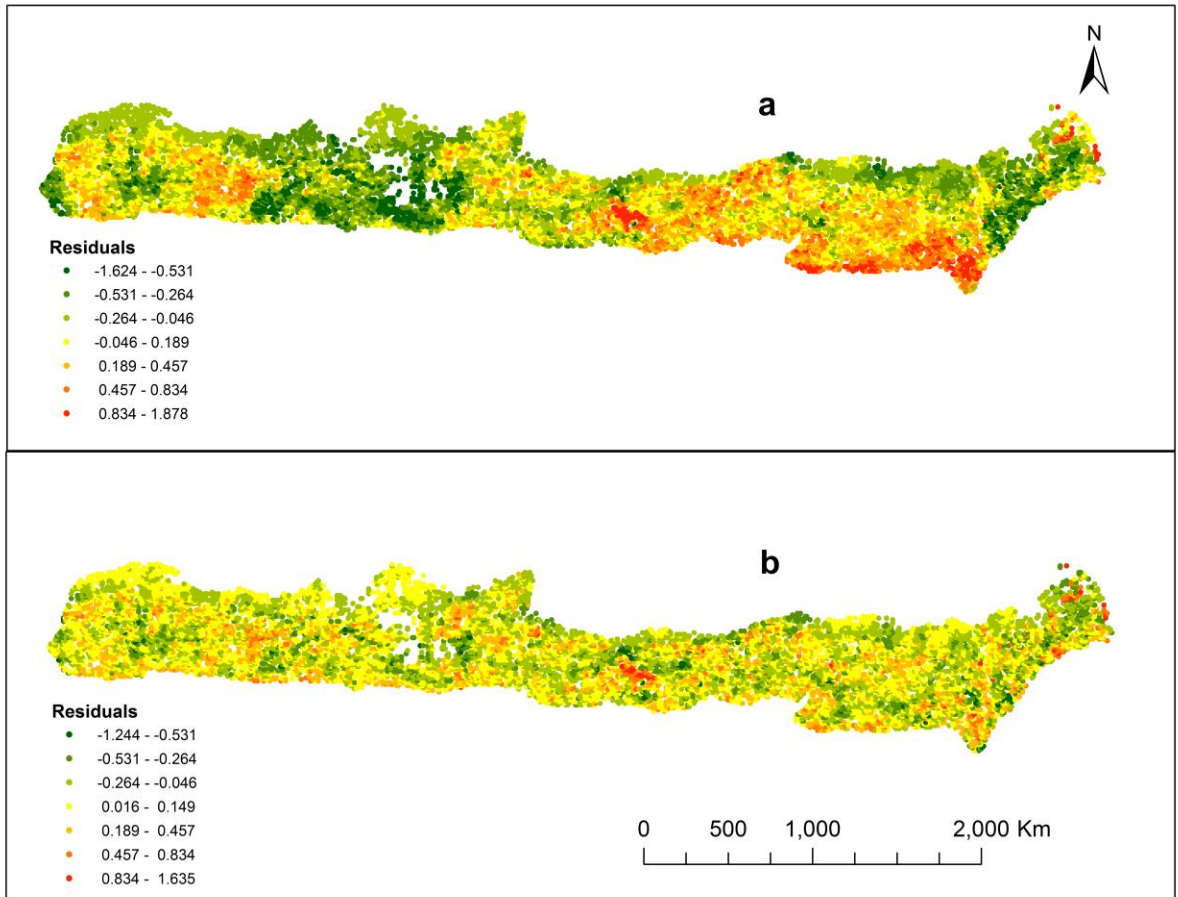
High amounts of residual autocorrelation usually indicate problems with model specification and non-stationarity. The violation of spatial independence can result in unstable regression results. In general, a random distribution of residuals suggests more reliable regression models. Figures 14 and 15 illustrate the residual distribution for both years for their corresponding OLS and GWR models, respectively. From the spatial patterns, it is clear that that the local modelling approach produces more randomized patterns. The spatial correlograms (Figure 16) that compute the Moran's I index at various spatial lags are more descriptive on how the regression models treats the residuals. Both the global OLS model and the OLS models run in separate land covers



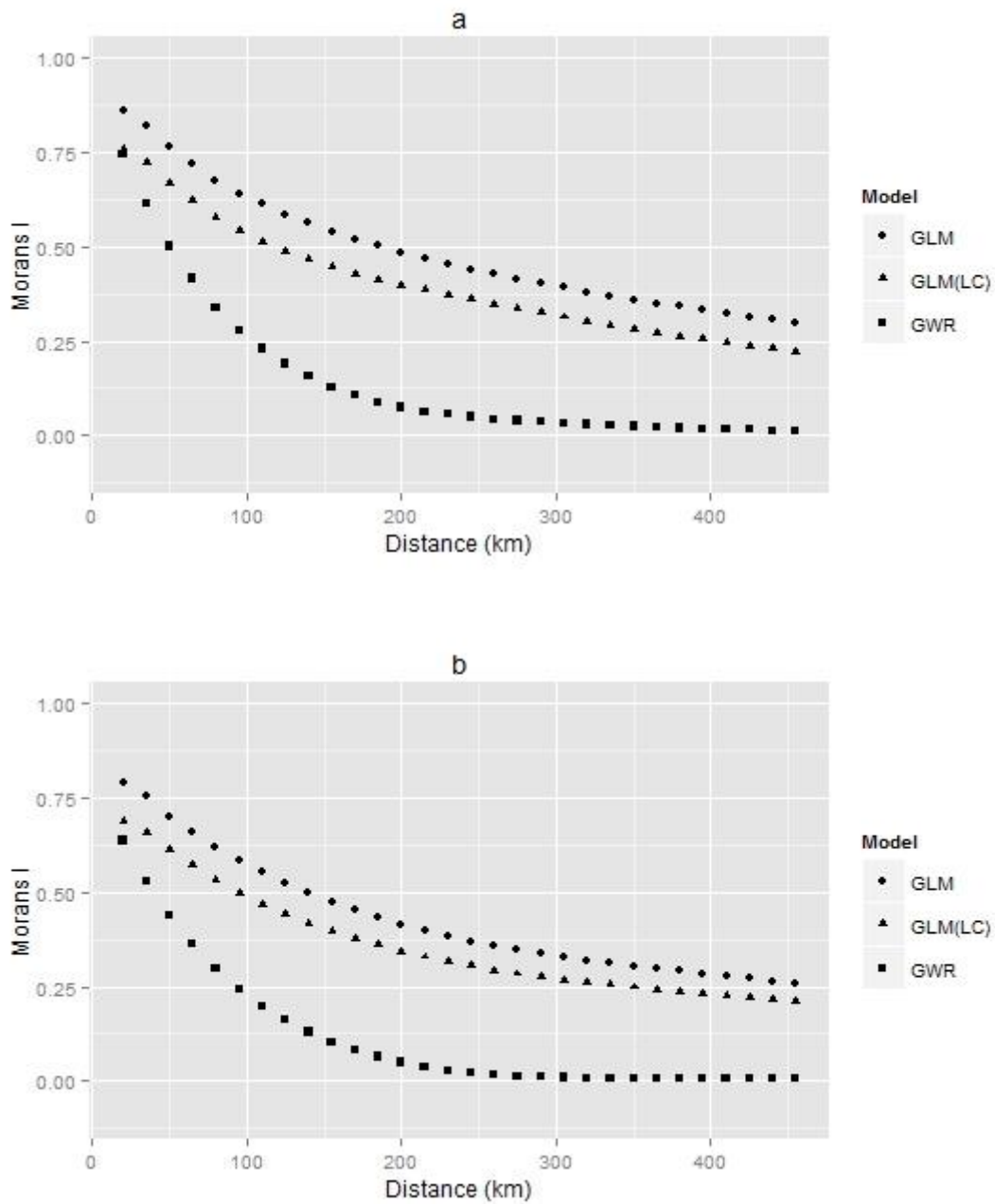
suffer from very high residual autocorrelation not only over short spatial lags, but also over very large distances. On the contrary, the GWR approach eliminated spatial autocorrelation at larger distances and significantly reduced residual dependence over short spatial lags. In all cases, the values of Moran's I were significant ( $p < 0.05$ ). This further promotes the use of spatial disaggregation modeling as it deals with residuals in a very direct manner.



**Figure 13.** Residual distribution in 2002 for OLS (a) and GWR (b) models.



**Figure 14.** Residual distribution in 2012 for OLS (a) and GWR (b) models.



**Figure 15.** Spatial correlograms in 2002 (a) and 2012 (b). Circles denote the global OLS model (GLM), Triangles denote the combined OLS models ran in separate land cover categories (GLM–LC) and squares refer to the GWR model. All Moran's I values were significant ( $p < 0.05$ ).



## **5. Discussion**

The results demonstrated that the NDVI–rainfall relationship is not stationary throughout the Sahel during the growing season in the years under examination. Therefore, it is not reasonable to describe the relationship as a set of spatially constant coefficients. The spatial heterogeneity of rainfall and NDVI was documented by Foody (2002) for the whole of northern Africa, of which the Sahel is a part of. In that study, the relationship was positive for the whole extent of the Sahel but this can be attributed to the different datasets used. The spatial variations were attributed to varying land covers, soil patterns, and the quality of the datasets or other undocumented factors. In this study, the relationship was positive for the majority of the study area, however, negative or very weak relationships were observed in some locations as well. Moreover, the significance of the relationship varied dramatically through space.

The gradual change from arid and semi-arid environments to more humid environments was one reason as shown by the different regression results based on individual land cover examination. However, even though the OLS regressions ran in separate land covers produced better results in comparison to the global model that includes the whole extent of the Sahel, it did not significantly reduce residual autocorrelation and the predictions were subpar in comparison to the GWR model. Using land cover maps to improve modelling was shown to be of limited help for many reasons. First, land cover maps vary depending on the producer as these are not standardized. Second, it would be rather unreasonable to assume that the same land cover in one part of the Sahelian belt would have the same attributes in distances as one several thousand kilometres away from it. Finally, land cover maps are heavily aggregated, and although very helpful for exploration and interpretation purposes, lack local information as the lack of reduction in regression residual autocorrelation has shown.

GWR transcended arbitrarily defined barriers of land covers that often suffer from spatial discontinuity in climatically heterogeneous areas. By incorporating local information, GWR worked as a continuously varying detector of geographical relationships and significantly improved over the global models, by explaining more variance and effectively reducing autocorrelation in the residuals. The reason is that local modelling

can potentially take into account within land cover variation, different species composition and distribution and factors that have a strong local component such as soil type or human disruption of ecological communities and locally unique climatic conditions (Propastin 2009). These local variations demonstrated that the spatial patterns of NDVI are better correlated with rainfall in some parts of the Sahel than others. In general, the more humid parts were less correlated; for example, the wetlands around Lake Chad, potentially due to the abundance of water, different irrigation practices and increased human activity. In addition, the rainfall threshold that some vegetation types need to achieve maximum growth such as crops can be trespassed and therefore producing very weak correlations and associations with rainfall in the growing season (Sun et al. 2010).

As an ecogeographical transition zone between arid and humid environments, prominent areas were documented that appeared very sensitive to variations in accumulated rainfall for the growing season. Zhao et al. (2015) showed that the transition area between desert and humid lands in Northern China appeared more sensitive than surrounding areas to climatic changes. The Sahel, instead of being a continuous sensitive zone, appears to have large clusters of high correlations and rainfall coefficients e.g. a large cluster covering and extending over Sudan.

Another interesting finding of the study is that the results depicted temporal variability in their spatial relations - 2002 was a relatively dry year, while 2012 can be described as a wet year. Although, the results suffer from a decade of temporal discontinuity, the assumption is that the amount of rainfall in each area received in a given period is critical for the output of the results. In 2002, higher local fits and rainfall coefficients were observed in comparison to 2012. When rainfall is restricted, its effect as a predictor variable and the sensitivity of vegetation to it appear to be increased more than during wetter years. However, in 2012 there were areas that produced very high slope coefficients that surpassed even the highest slope coefficients in 2002 (Sudanian zone). This could be attributed to land cover change, a difference in irrigation practices or as a concerning uncertainty in respect to the quality and validity of the data. Regardless, this is a testament of the ability of GWR to provide local areas of interests that can direct

future research and improve the modelling of phenomena in these regions. These temporal variations cannot be shown by any OLS or SAR model and can otherwise mislead the interpretation of the results.

The relationship between vegetation and climatic variables such as rainfall has been shown to be scale dependent based on previous research (Propastin et al. 2008; Zhao et al. 2015). Gao and Li (2011) emphasized the importance of selecting an appropriate bandwidth and spatial weighting method. Fixed weighting has more commonly been employed in ecological studies when there is a continuous and relatively dense distribution of sample points as more robust interpretations regarding the effect of scale on the results can be made. In the current study, a fixed weighting scheme was used when GWR was applied to the whole extent of the Sahel, while adaptive weighting was selected to calibrate GWR models in separate land covers to combat spatial discontinuity in the sampling. The selection of the bandwidth corresponds to the scale at which the data will be examined and is perhaps the most important calibration setting in GWR. The bandwidth choice is a tradeoff between variance in the local estimates and bias in the model. Usually, AICc is employed as a tool on which to take the decision. However, it has to be noted that with large sample sizes (thousands of samples) AICc can suggest optimality in models with extremely small bandwidths, inflated  $R^2$  and large standard errors, hindering meaningful inferences and inserting a large amount of noise in the outputs (Propastin et al. 2008). This has not been noted often enough in environmental studies that employ GWR and should be a cause for concern for those who employ these methods. Here, AICc was very useful making comparisons between an OLS and a GWR model after the appropriate bandwidth was selected based on the Stationarity Index (SI). The AICc as a guide for bandwidth selection, is not a panacea when it comes to GWR calibration, and is very likely to encounter strong collinearity problems in the local explanatory terms – even between the intercept and the beta coefficient. Although smaller bandwidths can be useful for exploratory purposes, it is strongly suggested that covariance matrices, local correlation coefficients and condition indexes be applied before accepting a bandwidth even when it is suggested by measures such as AICc. Nakaya (2001) employed stricter measures for bandwidth selection such as Bayesian

Information Criteria (BIC), which penalizes smaller bandwidths more and could be of potential use in further research in this sort of ecological analysis.

The computation of the Stationarity Index (SI) at various incrementing spatial scales provided important information with regard to the scale dependency of NDVI and accumulated rainfall in the growing season in the Sahel. In the Sahel, for both years under examination the relationship tended to stabilize at roughly 200 kilometres and the regressions constructed at that scale are more reliable and stable as the condition indices and local correlation coefficients did not suggest any spurious local correlation in the coefficients, even at the cost of introducing more bias than smaller bandwidths. In similar studies (Gao et al. 2012, Zhang et al. 2015) in transitional zones between arid to humid environments that used similar data, for example the Qinghai–Tibet and Mongolian plateaus, the stability scale was 120 and 500 kilometres respectively. Agreeing with the results of these studies, this implies that the operational scale between vegetation and climatic variables can vary regionally and is study area specific. Additionally, it has to be noted that in these studies annual comparisons were performed while in this study the focus was in the correlations in the growing season. By analysing the data only in the months where vegetation growth and most of the rainfall occurs much of the unwanted noise from the NDVI signal in the dry season can be eliminated. However, this methodology is strictly related to the climatic conditions of each study area.

The findings of this thesis can be used for modelling net primary product in the Sahel as a function of location as it was shown that rainfall and NDVI relations are spatially varying. Both NDVI and rainfall are often components in NPP models and their interactions can vary locally (Wang et al. 2005). Additionally, it could be employed as a method for estimating woody biomass assuming that a linear regression is not adequate enough to capture the true relationship in a heterogeneous environment as the Sahel. Lastly, GWR analysis by employing data from different satellite sources may increase our understanding regarding the effect of the spatial resolution of the data in the estimated relations of vegetation and climatic variables.

The interpretation of results is a topic of concern as Jetz et al. (2005) suggested that relationships in environmental variables are global and likely vary due to missing



variables or interaction terms. It is quite difficult to take a justifiable stand on that but the results of this thesis disagree with the conclusions of Jetz et al. (2005), and showed that relationships can and do in fact, vary in space. Moreover, an important decrease in the autocorrelation of the residuals was noticed which was a major part of the criticism of GWR in the aforementioned study. It would be more likely to assume that, incorrect statistical inferences can be concluded when we conduct a GWR analysis without using appropriate diagnostic tools to test for local collinearity and other problems, as described in detail by Wheeler and Tiefelsdorf (2007) and Paez et al. (2011). Moreover, the application of GWR has advanced in the last years and can be applied as a mixed model that assumes spatial constancy for one or more of the variables (Mei et al. 2006) or spatiotemporal models (Huang et al, 2010) that incorporate temporal information in their calibrations. These models can be very useful in areas such as the Sahel, which are sensitive to environmental changes both in the spatial and temporal domain.

The study demonstrated that GWR can be an alternative to OLS modelling in heterogeneous areas that are sensitive to environmental changes. The local method produced better predictions, lower autocorrelation in the residuals and highlighted interesting local variations. Another way to achieve similar results would be to divide the Sahel into geographical regions that share common attributes as Zhao et al. (2015) showed in their research in Northern China. However, that would be a very tedious process and might be untenable when it comes with a large amount of data from both a spatial and temporal perspective. Therefore, GWR is strongly suggested as both an explanatory and exploratory method in environmental modelling when spatial constancy in relations between variables is questionable.



## 6. Conclusions

In a large scale analysis, the NDVI-rainfall relationship often emerges with characteristics such as non-stationarity and scale dependency, especially in transition zones between different climatic environments. In this thesis, the spatial variation and scale dependency of that relationship was revealed in the Sahel of Africa, a prominent transition zone between arid to humid environments which have been in the spotlight of scientific research since the past decades due to its sensitive spatiotemporal dynamics. By taking into account the existence of different types of land cover, and the extreme heterogeneity in climatic conditions from the southern Sahara boundaries to the humid areas of the South, the effect of rainfall upon the spatial patterns of vegetation was the main interest of this study. The analysis was restricted to the growing season to avoid unnecessary noise coming from NDVI during the dry months where rainfall is essentially zero. Lastly, data from two years were used - 2002 and 2012 which correspond to relatively dry and wet years, respectively.

By undertaking a spatial disaggregation modelling technique named Geographically Weighted Regression (GWR) and after finding the most appropriate scale to examine the spatial relationship, the results highlighted areas which were particularly sensitive to variations in rainfall and which seemingly form large clusters that connect humid and arid climatic zones. In these areas, rainfall appears to be the dominant determinant in understanding the distribution of vegetation. Moreover, regions mainly located around wetlands were shown to have a very weak relationship with rainfall indicating the need for incorporating additional variables to explain the NDVI variation. Finally, temporal variations were showcased as the spatial relationships would often change from a drier year to a more humid one.

In comparison with traditional linear regression modelling such as Ordinary Least Squares (OLS), GWR model performed significantly better in both years, by producing more accurate predictions, reducing autocorrelation in the regression residuals and allowing for local inferences to be made due to a large output from GWR results being a set of maps showcasing the local situation between NDVI and rainfall. The results were validated by conventional regression diagnostics and local tests to assess the significant

and degree of non-stationarity in the data. Therefore, GWR is suggested as an accurate, informative technique both for exploratory and explanatory reasons to treat non-stationarity in heterogeneous areas in an ecological context.

## **Acknowledgements**

I would like to thank my supervisors Abdulhakim Abdi and David Tenenbaum for their valuable feedback and contributions and Dr. Kalogirou for his insights during the analysis. Moreover, my studies in Lund would not be possible without the help and support of my family. Finally, and in no particular order, I would like to thank my friends who made my stay here enjoyable and helped me through difficult times: Abdalla Eltayeb, Jayan sri Jeewantha Wajishingha, Christos Bountzouklis, Marie Baehr, Matt Corney, Caesar Alsamaray and Gea Hallen.



## References

1. Abdi, A. M., Seaquist, J., Tenenbaum, D. E., Eklundh, L., & Ardö, J. (2014). The supply and demand of net primary production in the Sahel. *Environmental Research Letters*, 9(9), 094003.
2. Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6), 716-723.
3. Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical analysis*, 20(1), 1-17.
4. Anselin, L. (2003). *GeoDa™ 0.9 user's guide*. Urbana, 51, 61801.
5. Anyamba, A., & Tucker, C. J. (2005). Analysis of Sahelian vegetation dynamics using NOAA-AVHRR NDVI data from 1981–2003. *Journal of Arid Environments*, 63(3), 596-614.
6. Brandt, M., Mbow, C., Diouf, A. A., Verger, A., Samimi, C., & Fensholt, R. (2015). Ground-and satellite-based evidence of the biophysical mechanisms behind the greening Sahel. *Global change biology*, 21(4), 1610-1620.
7. Brunson, C., Fotheringham, S., & Charlton, M. (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3), 431-443
8. Bolduc, D., Laferriere, R., & Santarossa, G. (1995). Spatial autoregressive error components in travel flow models: an application to aggregate mode choice(pp. 96-108). Springer Berlin Heidelberg
9. Cahill, M., & Mulligan, G. (2007). Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review*, 25(2), 174-193.
10. Chen, D., & Brutsaert, W. (1998). Satellite-sensed distribution and spatial patterns of vegetation parameters over a tallgrass prairie. *Journal of the atmospheric sciences*, 55(7), 1225-1238.
11. Chamaille-Jammes, S., Fritz, H., & Murindagomo, F. (2006). Spatial patterns of the NDVI–rainfall relationship at the seasonal and interannual time scales in

- an African savanna. *International Journal of Remote Sensing*, 27(23), 5185-5200.
12. Cliff, A., & Ord, K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical analysis*, 4(3), 267-284.
  13. Davenport, M. L., & Nicholson, S. E. (1993). On the relation between rainfall and the Normalized Difference Vegetation Index for diverse vegetation types in East Africa. *International Journal of Remote Sensing*, 14(12), 2369-2389.
  14. Eklundh, L. (1998). Estimating relations between AVHRR NDVI and rainfall in East Africa at 10-day and monthly time scales. *International Journal of Remote Sensing*, 19(3), 563-570.
  15. Eklundh, L., & Olsson, L. (2003). Vegetation index trends for the African Sahel 1982–1999. *Geophysical Research Letters*, 30(8).
  16. Fensholt, R., & Rasmussen, K. (2011). Analysis of trends in the Sahelian ‘rain-use efficiency’ using GIMMS NDVI, RFE and GPCP rainfall data. *Remote Sensing of Environment*, 115(2), 438-451.
  17. Fensholt, R., & Proud, S. R. (2012). Evaluation of earth observation based global long term vegetation trends—Comparing GIMMS and MODIS global NDVI time series. *Remote sensing of Environment*, 119, 131-147.
  18. Foody, G. M. (2003). Geographical weighting as a further refinement to regression modelling: an example focused on the NDVI–rainfall relationship. *Remote sensing of Environment*, 88(3), 283-293.
  19. Foody, G. M. (2004). Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecology and Biogeography*, 13(4), 315-320.
  20. Fotheringham, A. S., Charlton, M. E., & Brunson, C. (2001). Spatial variations in school performance: a local analysis using geographically weighted regression. *Geographical and Environmental Modelling*, 5(1), 43-66.
  21. Fotheringham, A. S., Brunson, C., & Charlton, M. (2003). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.



22. Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data*.
23. Gao, J., & Li, S. (2011). Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using geographically weighted regression. *Applied Geography*, 31(1), 292-302.
24. Gao, Y., Huang, J., Li, S., & Li, S. (2012). Spatial pattern of non-stationarity and scale-dependent relationships between NDVI and climatic factors—A case study in Qinghai-Tibet Plateau, China. *Ecological Indicators*, 20, 170-176.
25. Gaughan, A. E., & Waylen, P. R. (2012). Spatial and temporal precipitation variability in the Okavango–Kwando–Zambezi catchment, southern Africa. *Journal of Arid Environments*, 82, 19-30.
26. Grist, J., Nicholson, S. E., & Mpolokang, A. (1997). On the use of NDVI for estimating rainfall fields in the Kalahari of Botswana. *Journal of Arid Environments*, 35(2), 195-214.
27. Hickler, T., Eklundh, L., Seaquist, J. W., Smith, B., Ardö, J., Olsson, L., ... & Sjöström, M. (2005). Precipitation controls Sahel greening trend. *Geophysical Research Letters*, 32(21).
28. Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3), 383-401.
29. Hu, M., Li, Z., Wang, J., Jia, L., Liao, Y., Lai, S., ... & Yang, W. (2012). Determinants of the incidence of hand, foot and mouth disease in China using geographically weighted regression models. *PloS one*, 7(6), e38978.
30. Jamali, S., Seaquist, J., Eklundh, L., & Ardö, J. (2014). Automated mapping of vegetation trends with polynomials using NDVI imagery over the Sahel. *Remote Sensing of Environment*, 141, 79-89.
31. Jetz, W., Rahbek, C., & Lichstein, J. W. (2005). Local and global approaches to spatial data analysis in ecology. *Global Ecology and Biogeography*, 14(1), 97-98.

32. Kalogirou, S. (2003). The statistical analysis and modelling of internal migration flows within England and Wales.
33. Kalogirou, S., & Hatzichristos, T. (2007). A spatial modelling framework for income estimation. *Spatial Economic Analysis*, 2(3), 297-316.
34. Kalogirou, S. (2013). Testing geographically weighted multicollinearity diagnostics. GISRUK
35. Kalogirou, S., & Kalogirou, M. S. (2015). Package 'lctools'.
36. Kalogirou, S. (2015). Destination Choice of Athenians: An Application of Geographically Weighted Versions of Standard and Zero Inflated Poisson Spatial Interaction Models. *Geographical Analysis*.
37. Karen, K. K. (2008). *Encyclopedia of Geographic Information Science*.
38. Kaspersen, P. S., Fensholt, R., & Huber, S. (2011). A spatiotemporal analysis of climatic drivers for observed changes in Sahelian vegetation productivity (1982–2007). *International Journal of Geophysics*, 2011.
39. Kissling, W. D., & Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, 17(1), 59-71.
40. Lamprey, H. F. (1988). Report on the desert encroachment reconnaissance in northern Sudan: 21 October to 10 November 1975. *Desertification Control Bulletin*
41. Latham, J., Cumani, R., Rosati, I., & Bloise, M. (2014). Global land cover share (GLC-SHARE) database beta-release version 1.0-2014. FAO: Rome, Italy.
42. Li, B., Tao, S., & Dawson, R. W. (2002). Relations between AVHRR NDVI and ecoclimatic parameters in China. *International Journal of Remote Sensing*, 23(5), 989-999.
43. Li, J., Lewis, J., Rowland, J., Tappan, G., & Tieszen, L. L. (2004). Evaluation of land performance in Senegal using multi-temporal NDVI and rainfall series. *Journal of Arid Environments*, 59(3), 463-480
44. Malo, A. R., & Nicholson, S. E. (1990). A study of rainfall and vegetation dynamics in the African Sahel using normalized difference vegetation index. *Journal of Arid Environments*, 19(1), 1-24.

45. Maantay, J. A., & McLafferty, S. (Eds.). (2011). Geospatial analysis of environmental health (Vol. 4). Springer Science & Business Media.
46. Maselli, F., & Chiesi, M. (2006). Integration of multi-source NDVI data for the estimation of Mediterranean forest productivity. *International Journal of Remote Sensing*, 27(1), 55-72.
47. Mei, C. L., Wang, N., & Zhang, W. X. (2006). Testing the importance of the explanatory variables in a mixed geographically weighted regression model. *Environment and Planning A*, 38(3), 587-598.
48. Mensching, H. G. (1990). Desertification: environmental destruction in arid zones. Wissenschaftliche Buchgesellschaft.
49. Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2), 243-251.
50. Nakaya, T. (2001). Local spatial interaction modelling based on the geographically weighted regression approach. *GeoJournal*, 53(4), 347-358.
51. Nakaya, T., Fotheringham, A. S., Charlton, M., & Brunsdon, C. (2009). Semiparametric geographically weighted generalised linear modelling in GWR 4.0.
52. Nicholson, S. E., Davenport, M. L., & Malo, A. R. (1990). A comparison of the vegetation response to rainfall in the Sahel and East Africa, using normalized difference vegetation index from NOAA AVHRR. *Climatic change*, 17(2-3), 209-241.
53. Nicholson, S. E., Tucker, C. J., & Ba, M. B. (1998). Desertification, drought, and surface vegetation: An example from the West African Sahel. *Bulletin of the American Meteorological Society*, 79(5), 815.
54. Olsson, L., Eklundh, L., & Ardö, J. (2005). A recent greening of the Sahel—trends, patterns and potential causes. *Journal of Arid Environments*, 63(3), 556-566.
55. Osborne, P. E., & Suárez-Seoane, S. (2002). Should data be partitioned spatially before building large-scale distribution models?. *Ecological Modelling*, 157(2), 249-259.

56. Páez, A., Farber, S., & Wheeler, D. (2011). A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A*, 43(12), 2992-3010.
57. Potter, C. S., & Brooks, V. (1998). Global analysis of empirical relations between annual climate and seasonality of NDVI. *International Journal of Remote Sensing*, 19(15), 2921-2948
58. Propastin, P., Kappas, M., & Erasmi, S. (2008). Application of geographically weighted regression to investigate the impact of scale on prediction uncertainty by modelling relationship between vegetation and climate. *International journal of spatial data infrastructures research*, 3, 73-94
59. Propastin, P. A. (2009). Spatial non-stationarity and scale-dependency of prediction accuracy in the remote estimation of LAI over a tropical rainforest in Sulawesi, Indonesia. *Remote Sensing of Environment*, 113(10), 2234-2242.
60. Richard, Y., & Pocard, I. (1998). A statistical study of NDVI sensitivity to seasonal and interannual rainfall variations in Southern Africa. *International Journal of Remote Sensing*, 19(15), 2907-2920.
61. Santos, P., & Negri, A. J. (1997). A comparison of the normalized difference vegetation index and rainfall for the Amazon and Northeastern Brazil. *Journal of Applied Meteorology*, 36(7), 958-965.
62. Sun, H., Shen, Y., Yu, Q., Flerchinger, G. N., Zhang, Y., Liu, C., & Zhang, X. (2010). Effect of precipitation change on water balance and WUE of the winter wheat–summer maize rotation in the North China Plain. *Agricultural Water Management*, 97(8), 1139-1145.
63. Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2), 127-150.
64. Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., . & El Saleous, N. (2005). An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, 26(20), 4485-4498.

65. Wang, J., Price, K. P., & Rich, P. M. (2001). Spatial patterns of NDVI in response to precipitation and temperature in the central Great Plains. *International Journal of Remote Sensing*, 22(18), 3827-3844.
66. Wang, K., Zhang, C., & Li, W. (2013). Predictive mapping of soil total nitrogen at a regional scale: a comparison between geographically weighted regression and cokriging. *Applied Geography*, 42, 73-85.
67. Wang, Q., Ni, J., & Tenhunen, J. (2005). Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. *Global Ecology and Biogeography*, 14(4), 379-393.
68. Wheeler, D., & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7(2), 161-187.
69. Zhang, X., Hu, Y., Zhuang, D., Qi, Y., & Ma, X. (2009). NDVI spatial pattern and its differentiation on the Mongolian Plateau. *Journal of geographical Sciences*, 19(4), 403-415.
70. Zhao, N., Yang, Y., & Zhou, X. (2010). Application of geographically weighted regression in estimating the effect of climate and site conditions on vegetation distribution in Haihe Catchment, China. *Plant Ecology*, 209(2), 349-359.
71. Zhao, Z., Gao, J., Wang, Y., Liu, J., & Li, S. (2015). Exploring spatially variable relationships between NDVI and climatic factors in a transition zone using geographically weighted regression. *Theoretical and Applied Climatology*, 120(3-4), 507-519.

## List of Publications

### Institutionen för naturgeografi och ekosystemvetenskap, Lunds Universitet.

Student examensarbete (Seminarieuppsatser). Uppsatserna finns tillgängliga på institutionens geobibliotek, Sölvegatan 12, 223 62 LUND. Serien startade 1985. Hela listan och själva uppsatserna är även tillgängliga på LUP student papers (<https://lup.lub.lu.se/student-papers/search/>) och via Geobiblioteket ([www.geobib.lu.se](http://www.geobib.lu.se))

The student thesis reports are available at the Geo-Library, Department of Physical Geography and Ecosystem Science, University of Lund, Sölvegatan 12, S-223 62 Lund, Sweden. Report series started 1985. The complete list and electronic versions are also electronic available at the LUP student papers (<https://lup.lub.lu.se/student-papers/search/>) and through the Geo-library ([www.geobib.lu.se](http://www.geobib.lu.se))

- 350 Mihaela – Mariana Tudoran (2015) Occurrences of insect outbreaks in Sweden in relation to climatic parameters since 1850
- 351 Maria Gatzouras (2015) Assessment of trampling impact in Icelandic natural areas in experimental plots with focus on image analysis of digital photographs
- 352 Gustav Wallner (2015) Estimating and evaluating GPP in the Sahel using MSG/SEVIRI and MODIS satellite data
- 353 Luisa Teixeira (2015) Exploring the relationships between biodiversity and benthic habitat in the Primeiras and Segundas Protected Area, Mozambique
- 354 Iris Behrens & Linn Gardell (2015) Water quality in Apac-, Mbale- & Lira district, Uganda - A field study evaluating problems and suitable solutions
- 355 Viktoria Björklund (2015) Water quality in rivers affected by urbanization: A Case Study in Minas Gerais, Brazil
- 356 Tara Mellquist (2015) Hållbar dagvattenhantering i Stockholms stad - En riskhanteringsanalys med avseende på långsiktig hållbarhet av Stockholms stads dagvattenhantering i urban miljö
- 357 Jenny Hansson (2015) Trafikrelaterade luftföroreningar vid förskolor – En

- studie om kvävedioxidhalter vid förskolor i Malmö
- 358 Laura Reinelt (2015) Modelling vegetation dynamics and carbon fluxes in a high Arctic mire
- 359 Emelie Linnéa Graham (2015) Atmospheric reactivity of cyclic ethers of relevance to biofuel combustion
- 360 Filippo Gualla (2015) Sun position and PV panels: a model to determine the best orientation
- 361 Joakim Lindberg (2015) Locating potential flood areas in an urban environment using remote sensing and GIS, case study Lund, Sweden
- 362 Georgios-Konstantinos Lagkas (2015) Analysis of NDVI variation and snowmelt around Zackenberg station, Greenland with comparison of ground data and remote sensing.
- 363 Carlos Arellano (2015) Production and Biodegradability of Dissolved Organic Carbon from Different Litter Sources
- 364 Sofia Valentin (2015) Do-It-Yourself Helium Balloon Aerial Photography - Developing a method in an agroforestry plantation, Lao PDR
- 365 Shirin Danehpash (2015) Evaluation of Standards and Techniques for Retrieval of Geospatial Raster Data - A study for the ICOS Carbon Portal
- 366 Linnea Jonsson (2015) Evaluation of pixel based and object based classification methods for land cover mapping with high spatial resolution satellite imagery, in the Amazonas, Brazil.
- 367 Johan Westin (2015) Quantification of a continuous-cover forest in Sweden using remote sensing techniques
- 368 Dahlia Mudzaffar Ali (2015) Quantifying Terrain Factor Using GIS Applications for Real Estate Property Valuation
- 369 Ulrika Belsing (2015) The survival of moth larvae feeding on different plant species in northern Fennoscandia
- 370 Isabella Grönfeldt (2015) Snow and sea ice temperature profiles from satellite data and ice mass balance buoys
- 371 Karolina D. Pantazatou (2015) Issues of Geographic Context Variable Calculation Methods applied at different Geographic Levels in Spatial

- Historical Demographic Research -A case study over four parishes in Southern Sweden
- 372 Andreas Dahlbom (2016) The impact of permafrost degradation on methane fluxes - a field study in Abisko
- 373 Hanna Modin (2016) Higher temperatures increase nutrient availability in the High Arctic, causing elevated competitive pressure and a decline in *Papaver radicum*
- 374 Elsa Lindevall (2016) Assessment of the relationship between the Photochemical Reflectance Index and Light Use Efficiency: A study of its seasonal and diurnal variation in a sub-arctic birch forest, Abisko, Sweden
- 375 Henrik Hagelin and Matthieu Cluzel (2016) Applying FARSITE and Prometheus on the Västmanland Fire, Sweden (2014): Fire Growth Simulation as a Measure Against Forest Fire Spread – A Model Suitability Study –
- 376 Pontus Cederholm (2016) Californian Drought: The Processes and Factors Controlling the 2011-2016 Drought and Winter Precipitation in California
- 377 Johannes Loer (2016) Modelling nitrogen balance in two Southern Swedish spruce plantations
- 378 Hanna Angel (2016) Water and carbon footprints of mining and producing Cu, Mg and Zn: A comparative study of primary and secondary sources
- 379 Gusten Brodin (2016) Organic farming's role in adaptation to and mitigation of climate change - an overview of ecological resilience and a model case study
- 380 Verånika Trollblad (2016) Odling av *Cucumis Sativus* L. med aska från träd som näringstillägg i ett urinbaserat hydroponiskt system
- 381 Susanne De Bourg (2016) Tillväxteffekter för andra generationens granskog efter tidigare genomförd kalkning
- 382 Katarina Crafoord (2016) Placering av energiskog i Sverige - en GIS analys
- 383 Simon Nåfält (2016) Assessing avalanche risk by terrain analysis An experimental GIS-approach to The Avalanche Terrain Exposure Scale (ATES)
- 384 Vide Hellgren (2016) Asteroid Mining - A Review of Methods and Aspects
- 385 Tina Truedsson (2016) Hur påverkar snömängd och vindförhållande



- vattentrycksmätningar vintertid i en sjö på västra Grönland?
- 386 Chloe Näslund (2016) Prompt Pediatric Care Pediatric patients' estimated travel times to surgically-equipped hospitals in Sweden's Scania County
- 387 Yufei Wei (2016) Developing a web-based system to visualize vegetation trends by a nonlinear regression algorithm
- 388 Greta Wistrand (2016) Investigating the potential of object-based image analysis to identify tree avenues in high resolution aerial imagery and lidar data
- 389 Jessica Ahlgren (2016) Development of a Web Mapping Application for grazing resource information in Kordofan, Sudan, by downloading MODIS data automatically via Python
- 390 Hanna Axén (2016) Methane flux measurements with low-cost solid state sensors in Kobbefjord, West Greenland
- 391 Ludvig Forslund (2016) Development of methods for flood analysis and response in a Web-GIS for disaster management
- 392 Shuzhi Dong (2016) Comparisons between different multi-criteria decision analysis techniques for disease susceptibility mapping
- 393 Thirze Hermans (2016) Modelling grain surplus/deficit in Cameroon for 2030
- 394 Stefanos Georganos (2016) Exploring the spatial relationship between NDVI and rainfall in the semi-arid Sahel using geographically weighted regression
- 395 Julia Kelly (2016) Physiological responses to drought in healthy and stressed trees: a comparison of four species in Oregon, USA

